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

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

Apr. 2022 Journal of Educational Technology & Society vol.25 no.2

Journal of Educational Technology & Society

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

Volume 25 Issue 2
April 2022
ISSN: 1436-4522 (online)
ISSN: 1176-3647 (print)
http://www.j-ets.net/

Educational Technology & Society

An International Journal

Aims and Scope

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

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

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

Publisher

International Forum of Educational Technology & Society

Host

National Taiwan Normal University, Taiwan

Editorial Office

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

Supporting Organizations

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

Website Maintenance

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

Advertisements

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

Abstracting and Indexing

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

Guidelines for authors

Submissions are invited in the following categories:

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

All peer review publications will be 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. Please place figures in the manuscript.

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

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

Educational Technology & Society

Volume 25 Number 2 2022

Table of Contents

Full Length Articles

Editorial Note

Special Issue Articles

A Programming Disposition Scale for High School Students

Chiu-Fan Hu[1,3], Yu-Tzu Lin[1,3], Cheng-Chih Wu[1,3*] and Hsueh-Chih Chen[2,3]
[1]Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan //
[2]Department of Educational Psychology and Counseling, National Taiwan Normal University, Taiwan //
[3]Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan //
chiufan@ntnu.edu.tw // linyt@ntnu.edu.tw // chihwu@ntnu.edu.tw // chcjyh@ntnu.edu.tw
[*]Corresponding author

(Submitted March 25, 2021; Revised September 3, 2021; Accepted September 4, 2021)

ABSTRACT: This study developed a scale to assess high school students' programming disposition. The scale was developed by utilizing a standardized test development process. The three constructs of the scale, namely confidence, persistence and flexible thinking, consisted of 9 items (3 items on each construct). Participants for the formal test of the scale were 1,332 students from 11 high schools. The validity and reliability of the programming disposition scale were validated via internal consistency, test-retest reliability, construct validity, discriminant validity, criterion-related validity, correlation coefficient of each subscale and confirmatory factor analysis. The analysis results showed that this scale is valid and reliable. The scale can serve as an assessment tool to assist teachers to instruct students learning programming, and help students determine whether taking programming courses in high school or pursuing programming-related majors in university. The effects of individual differences on programming disposition were also discussed to provide feasible educational implications.

Keywords: Disposition, Programming, Assessment tool, High school students

1. Introduction

A disposition is a tendency to display particular behaviors in a certain situation (Biber et al., 2013). Various patterns of thinking, such as confidence and attitude, enable one to be critical, thoughtful, and willing to work in a complex society (Wilkins, 2000). It includes not only students' confidence, curiosity, values and attitudes but also flexible thinking and the development of strategies for problem solving (Whitin, 2007). Students' inclinations and dispositions serve as predictors for their likelihood of taking related courses and pursuing various fields of study (Wilkins, 2000). The importance of student inclination and tendencies (disposition) has been previously addressed in the area of mathematics education. The National Council of Teachers of Mathematics (NCTM, 1989; NCTM, 2000) repeatedly stated the importance for teachers to improve and assess students' mathematical disposition. A positive disposition towards mathematics is considered to be more important than mathematical knowledge (Kusmaryono et al., 2019; Wilkins, 2000).

Programming is a subject related to mathematics and is considered as an integral component of K-12 curriculum as mathematics in many countries as it is a systematic way of approaching problem solving (Burrus & Moore, 2016; Winslow, 1996). In fact, programming has become an essential subject in K-12 schools to cope with the need of learning computational thinking (Lye & Koh, 2014). However, high school students often feel frustrated in learning text-based programming and have lower learning motivation (Galgouranas & Xinogalos, 2018). This would also affect students' academic intention for advanced study (Grandell et al., 2005). It might be beneficial if we can promote students' programming disposition. However, there is still less relevant research.

The development of assessment tools for disposition is still an open problem. To assess one's mathematical disposition, several tools have been developed to identify students' beliefs and attitudes (Royster et al., 1999), confidence (Wilkins, 2000), persistence (Breen et al., 2010), and flexible thinking (Whitin, 2007). Regarding programming, there is no assessment tool for disposition. In fact, it seems more challenging to develop a valid tool for assessing programming disposition because programming involves more various knowledge (e.g., programming syntax, constructs, and computer architecture) and skills (e.g., the use of IDEs, coding, and debugging). As Tsai et al. (2019) indicated, there is a lack of assessment tools for programming disposition in high school.

To fill research gaps, this study aims to develop a standardized scale to assess high school students' programming disposition. The disposition was assessed in terms of students' confidence, persistence, and flexible thinking on learning programming. Accordingly, the following research questions were explored:

Q1. Is the proposed programming disposition scale a valid and reliable assessment tool?
Q2. Does the second-order model of programming disposition show a good goodness of fit?

2. Literature review

2.1. Programming learning

Learning programming skills is often seen as difficult (Fitzgerald et al., 2008; Rum & Ismail, 2017; Sáez-López et al., 2016). The difficulties often deal with the abstract nature of programming (Bennedsen & Caspersen, 2006), intensive problem solving (Yurdugül & Aşkar, 2013), and using complex hierarchy of skills (Gray et al., 1993). Many studies have discussed that reducing the learning difficulties may also be linked with various attitudinal issues rather than intrinsic complexity of programming, such as complexity of syntax and algorithms (Hu et al., 2020; Luxton-Reilly, 2016). The idea of supporting and developing positive attitude in students has received considerable attention in programming education. Hu et al. (2021) advocated that programming instruction should emphasize arousing students' interests and improving attitudes rather than developing complex knowledge and skills only. Previous research has suggested that it is essential to develop K-12 students' dispositions in a curriculum (Katz, 1993). Students with a positive disposition have a curiosity in learning, appreciate the usefulness of learning subjects, are more confident of problem solving, and consequently, they are more disposed to apply their ability (Kusmaryono et al., 2019). What is more, without proper instruction to arouse students' disposition, students might have a negative disposition in learning. Students' attitudes towards programming have been investigated from various perspectives, such as self-efficacy (Sun & Hsu, 2019; Tsai et al., 2019), confidence in programming skills (Eliasson et al., 2006), and persistence of long-term learning (Eliasson et al., 2006; Gomes et al., 2012). However, there are few studies targeted on investigating students' programming dispositions. In addition, computer science educators are concerned about the lack of readily available, validated, or standardized assessment instruments in the field (Margulieux et al., 2019; Tew & Dorn, 2013). A rigorous process to develop the instruments is needed.

Besides programming disposition, there are still other factors that affect students' learning of programming, such as gender (Baser, 2013; Kong et al., 2018; Master et al., 2016), mathematical skills and abilities (Burrus & Moore, 2016; Erümit, 2020), science learning (Durak & Saritepeci, 2018), and parental support (Mason & Rich, 2020; Master et al., 2017). These factors might also affect students' programming disposition.

2.2. Construct of programming disposition

Student's programming abilities are correlated with their mathematical skills (Byrne & Lyons, 2001). The training of logical and abstract thinking, and reasoning in mathematics are relevant to working with abstract concepts and symbol manipulation in programming (Pioro, 2006). Students' mathematical dispositions served as a major foundation and springboard in our developing the construct of programming disposition. The NCTM (2000, see Table 1) has described students' dispositions as being relevant to their efforts in solving difficult problems and observing complex patterns, regularities, and correlations; these dispositions include confidence, perseverance, flexible thinking, and curiosity (NCTM, 2000; Whitin, 2007). Programming has been found as an effective tool for practicing computational thinking (Grover & Pea, 2013). The disposition towards computational thinking proposed by International Society for Technology in Education and the Computer Science Teachers Association (ISTE & CSTA, 2011, see Table 1) is also included as an important reference. The reference of NCTM and ISTE/CSTA constructs, along with literature in learning programming, allowed the construction of programming disposition scale to focus upon confidence, perseverance, and flexible thinking. The arguments are provided below.

Table 1. Constructs of mathematics/computational-thinking disposition

Mathematics disposition NCTM (2000)	Computational thinking disposition ISTE/CSTA (2011)
Confidence	Confidence in dealing with complexity
Perseverance	Persistence in working with difficult problems
Flexible thinking	Tolerance for ambiguity
Curiosity	Ability to deal with open-ended problems
	Ability to communicate and work with others to achieve the goal

Students' confidence and persistence (or perseverance) are both identified by NCTM and ISTE/CSTA as being important factors. Individual's confidence in dealing with complex problems is an important personal trait for

learning computer programming. Golding's et al. (2006) study has found that confidence was the most significant factor affecting one's performance in learning programming. There was a significantly positive correlation between students' confidence and their achievements in learning programming (Anastasiadou & Karakos, 2011; Baser, 2013). A student's level of confidence was found to be a major factor involved with the mastery of programming and especially for novices when trying to solve a complex problem (Eliasson et al., 2006).

Persistence, in terms of educational research, has been explained by many as a kind of continuously learning--one's tendency to pursue academic objectives (Pérez, 2018). In programming, persistence refers to continuing engagement when performing a challenging task. Persistency is needed to become a good programmer (Cheah, 2020; Jiau et al., 2009). Charlton and Birkett (1999) revealed that persistence is a predictor of programming achievement. Gomes et al. (2012) found persistence as being the most important reason students increase their performance in a programming course. Katz et al. (2006) also have found that students' persistence in programming correlated strongly with their grades. Perseverance (delineated by NCTM) has a very similar meaning with persistence applies to success in tackling difficult problems.

Flexible thinking has been characterized as the ability to restructure and transfer one's knowledge; that is, it enables people to understand, negotiate, and balance diverse views and beliefs-- those used to reach workable solutions (Barak & Levenberg, 2016). The process of learning programming does, indeed, involve such flexible thinking (Jang & Lew, 2014). One's personal flexibility is also an important characteristic in programming, such as approaching problems in multiple ways, being open to new ideas, and being open-minded (Begel & Nagappan, 2008). Concepts of flexible thinking include the disposition towards the following: "reflectivity, willingness to consider evidence contradictory to beliefs, willingness to consider alternative opinions and explanations, and a tolerance for ambiguity." This is also combined with a willingness to postpone closure (Stanovich & West, 1997). In this regard, the 'tolerance for ambiguity' is addressed in ISTE/CSTA and is a critical component of flexible thinking. The 'curiosity' delineated by NCTM is also a factor involved with flexible thinking. Students' exploratory attitudes and interests often manifest themselves with increased confidence while displaying flexibility and adaptability (Stokoe, 2012). These are aligned to concepts involved with flexible thinking.

The constructs relevant to 'ability' proposed by ISTE/CSTA were, additionally, removed because we focused on exploring students' programming dispositions (habits of mind) rather than their abilities (capabilities of doing something with knowledge and skills). Consequently, the scale utilized here consisted of three major constructs: confidence, persistence, and flexible thinking.

3. Method

We applied the standardized test development process to the development of the programming disposition scale used in this study. This development process involved two phases: (1) a pilot study and (2) a formal test. The pilot study was used to generate and analyze items. The formal test was used to examine the reliability and validity of the scale.

3.1. Participants

In the pilot study, convenience sampling was used to select 246 students (who did not participate in the formal test) from grades 10 to 12 who had learning experiences in programming from four Taipei high schools. In the formal study, the sample consisted of 117 (48%) tenth-grade, 76 (31%) eleventh-grade and 53 (22%) twelfth-grade students.

Table 2. The distribution of samples by school, grade, and academic track

Academic track	10^{th} grades	11^{th} grades		12^{th} grades		Total
	-	Science	Social science	Science	Social science	
Schools						
Tier 1	345	101	83	118	137	784
Tier 2	241	91	36	127	53	548
Total	586	192	119	245	190	1,332
		311		435		

Participants for the formal test of this study consisted of 1,332 students from 11 high schools in the Taipei metropolitan area in Taiwan. Stratified sampling was applied when recruiting the students. First, high schools were divided into two groups, Tier 1 and Tier 2, according to their traditional academic performance. Five to six schools were selected from each school group. Second, each school recruited one or two classes of students from each of the 10th to 12th grades. Finally, for 11th and 12th grades, both science and social science track students were recruited. High school students in Taiwan were divided into the two academic tracks after the 10th grade for their subject study. The distribution of samples by schools, grade, and academic track is shown in Table 2. All participants have programming experience because programming is covered in the 10th grade curriculum.

3.2. Procedure

The programming disposition scale was conducted on students in the formal test either by paper-and-pencil (two schools) or online (nine schools). The time for students to take the test was approximately 15 to 20 minutes.

3.3. Instruments

The programming disposition scale used here was developed based upon ones proposed by NCTM and ISTE/CSTA. The unique characteristics utilized in programming were considered when generating the constructs as discussed in section 2.2.

In the pilot study, draft items were adapted from various studies, such as "confidence" from the Fennema-Sherman Mathematics Attitudes Scales (Fennema & Sherman, 1976), "persistence" from (Breen et al., 2010), and "flexible thinking" from (Stanovich & West, 1997). Some items specifically related to programming aspects were added by the expert panel. A panel of seven experts included five computer science educators and two psychological and educational test professionals. They discussed and finalized 19 draft items (see Table 4) for further item analysis in the pilot study. Finally, a total of nine items were selected for the final scale used in the formal test. Three items were selected for each subscale (see Table 3). Item 6 is a negatively worded question which was reversed scored. The items developed here were selected based upon existing research, in which the scales used were mainly 5-point scales. Research by Croasmun and Ostrom (2011) has shown that a scale is both reliable and stable for both 4-point Likert and 5-point Likert scales. A 5-point Likert scale ranging from 1(strongly disagree) to 5 (strongly agree) was, thus, used in this study.

Table 3. Items of programming disposition scale

Constructs	Definition	Items
Confidence	Degree of having trust in programming	C1 I can get good grades in programming. C2 I can solve difficult programming tasks. C3 I believe I can learn programming.
Persistence	Continuing engagement in programming when facing a challenging task or spending a longtime to solve the task	P1 When presented with a difficult programming task, I increase my efforts. P2 I continue to work on a programming task even I have spent a long time to solve it and was not successful. P3 After learning programming for a while, I tend to give up.
Flexible thinking	Attempting to think differently or considering alternative solutions	FT1 I would try alternative solutions when solving problems similar to a previous one. FT2 I understand some programming tasks just cannot be solved in a short time. FT3 I consider alternative solutions when solving programming tasks.

Two instruments were used in this study to ensure the validity of the programming disposition scale. The Bebras Challenge (see https://www.bebras.org/) had over 2,872,000 students in 43 countries participated in 2019. The main goal of it is "to motivate pupils to be interested in informatics topics and to promote thinking which is algorithmic, logical, operational, and based on informatics fundamentals" (Dagienė & Stupuriene, 2016). The Bebras Challenge score was used to evaluate the correlation to the programming disposition scale in this study. The Comprehensive Assessment Program [CAP] for junior high school students is an examination for all 9th students in Taiwan. The examination scores play an important part for admitting students into secondary schools. CAP consists of Chinese, English, mathematics, natural science and social studies. This study used the CAP scores of mathematics and Chinese to assess the discriminant validity of programming disposition scale.

Additionally, three pieces of background information were collected from students, including gender (male, female), academic track (science, social science), and attitudes towards the degree of parental support (5-point Likert scale ranging from 1 to 5). This information was used to examine the construct validity of the programming disposition scale.

3.4. Data analysis

In the data analysis procedure, we analyzed data with SPSS 23.0 for Windows and LISREL 8.7 for Windows. Descriptive statistics were firstly performed to calculate the means, standard deviations and percentiles of student's programming disposition scores. Then, to test our research questions, the validity and reliability of this scale were evaluated using t tests and person correlation analysis to establish the internal consistency, test-retest reliability, criterion validity, discriminant validity and construct validity. A confirmatory factor analysis [CFA] was performed to identify the factor structure and items of the programming disposition scale. Independent t tests were used to examine the difference in gender and academic track. Pearson correlation analysis was conducted to test the correlations between parental support and programming disposition.

4. Results and discussion

4.1. Pilot study: Item analysis

The standards of evaluating included an improvement of internal consistency, item discrimination, factor loading, item-total correlation and individual item reliability. CFA results showed $\chi^2 = 769.18$ ($df = 149$), $p < .001$ and analysis of 19 items showed in Table 4.

Table 4. CFA results of 19 items

Construct	Item		Alpha if item deleted	Factor loading	Item-total correlation	Individual item reliability	*t*
Confidence	1.	I feel confident in programming.	.93	.85	.80**	.72	16.46***
	2.	I can get good grades in programming.[a]	.93	.79	.75**	.62	14.13***
	3.	I believe I can learn programming.[a]	.93	.82	.81**	.67	15.12***
	4.	I can solve difficult programming tasks.[a]	.93	.85	.82**	.72	16.33***
	5.	I cannot be good in programming.[b]	.93	.63	.62**	.40	11.43***
	6.	Programming is my worst learning activity.[b]	.93	.53	.55**	.28	9.62***
Persistence	7.	When presented with a difficult programming task, I increase my efforts.[a]	.93	.84	.83**	.71	16.34***
	8.	I tend to give up after spending much time on a programming task.[b]	.94	.32	.38**	.10	5.18***
	9.	I continue to work on a programming task even I have spent a long time to solve it and was not successful.[a]	.93	.85	.84**	.73	15.71***
	10.	I commit to spend a longtime to learn programming.	.93	.82	.81**	.67	17.03***
	11.	I believe learning programming requires a longtime effort.	.93	.46	.50**	.21	7.12***
	12.	After learning programming for a while, I tend to give up.[a b]	.93	.47	.54**	.22	9.29***
Flexible thinking	13.	I would try alternative solutions when I encountered difficulty in solving a programming task.	.93	.87	.83**	.76	17.37***
	14.	I always formulate solutions clearly before jumping into coding.	.93	.61	.63**	.37	9.47***
	15.	I would try alternative solutions when solving problems similar to a previous one.[a]	.93	.84	.77**	.70	14.35***

16. I understand some programming tasks just cannot be solved in a short time.[a]	.93	.60	.60**	.36	9.48***
17. I consider alternative solutions when solving programming tasks.[a]	.93	.84	.78**	.71	14.32***
18. I try to find out other solutions if I cannot solve a programming task.	.93	.82	.77**	.67	13.60***
19. I understand that not all problems can be solved by programming.	.94	.14	.13*	.02	1.78

Note. $^{*}p < .05$. $^{**}p < 0.01$. $^{***}p < 0.001$.[a] The item was included in the final programming disposition scale. [b] The item was a negative item.

First, according to the values of alpha if item deleted, each item was reliable (whole scale α = .93). The t-tests results of high and low scoring groups showed items had high discrimination (excluding item 19). The factors loaded between .14 and .87. Item 8 and 19 factor loading < .45. Further, the individual item reliability was between 0.02 and 0.76. 8 items (item 5, 6, 8, 11, 12, 14, 16 and 19) were considered to be deleted (individual item reliability < 0.5). The results of Pearson correlation showed that a significant correlation between each item and whole scale.

According to the results, item 1, 2, 3, 4, 7, 9, 15, 17 were included. In this scale, persistence means continuing engagement in programming when facing a challenging task or learning for a while. Compare to the other items, item 12 clearly states "after learning programming for a while," which could reflect the point in the persistence concept, "continuously for a while." As a result, we selected item 12 in the item pool. The concepts of flexible thinking include attempting to think carefully, considering alternative solutions and having a tolerance for ambiguity. The statement in item 16, "some programming tasks could not be solved soon" means that subjects needed to think more carefully or consider other possibilities, which was a kind of ambiguity. So item 16 was included. Finally, there were three items for each subscale. In the confidence subscale, item 1 to 4 were suggested to be included. However, concepts contained in item 2 to 4 already were enough to reflect item 1, in addition, to ensure the consistency in three subscales, we deleted item 1. Finally, the programming disposition scale was composed of 9 items.

According to the results of item analysis, the values of the goodness of fit were examined. The results found that $\chi^2 = 60.25$ (df = 24), $p < .001$, GFI = .95, AGFI = .90, RMR = .04, RMSEA = .07, NFI = .98, RFI = .96, CFI = .98, PGFI = .52, PNFI = .65, CN = 160.59. The results showed that the values of the goodness of fit are good.

4.2. Reliability and validity (Q1)

Table 5 shows the descriptive statistics of student's programming disposition in the formal test. The mean score for all participants was 28.45, averaged 3.22 for each item. Overall, students' programming disposition was found to be "medium" to "high." Students displayed the highest scores in flexible thinking (M = 10.08). Intermediate was that of persistence (M = 9.37), while confidence (M = 8.97) was shown to be the lowest. Our result with regard to "confidence" was similar to the TIMSS (Trends in International Science and Mathematics, 2020) study which showed that Taiwanese students lacked confidence in science and mathematics, although their performance has been shown to be higher than most of the countries (TIMSS, 2020).

Table 5. Descriptive statistics of student's programming disposition (N = 1,332)

		Total	Confidence	Persistence	Flexible thinking
M		28.42	8.97	9.37	10.08
SD		6.57	2.49	2.43	2.34
Min		9.00	3.00	3.00	3.00
Max		45.00	15.00	15.00	15.00
Percentiles	10	20.00	6.00	6.00	7.00
	20	23.00	7.00	7.60	8.00
	30	25.00	8.00	8.00	9.00
	40	28.00	9.00	9.00	10.00
	50	29.00	9.00	10.00	10.00
	60	30.00	10.00	10.00	11.00
	70	32.00	10.00	11.00	12.00
	80	34.00	11.00	11.00	12.00
	90	36.00	12.00	12.00	13.00

Cronbach's coefficient alpha (α) was used to test the internal consistency of the scale. The Cronbach's α of the entire scale was found to be .91. The subscales for confidence, persistence and flexible thinking were found to be .83, .78, and .78 respectively. The correlation coefficient for test–retest reliability was found to be .89 for the scale, and .86, .77, and .77 for the subscales of confidence, persistence, and flexible thinking, respectively. The correlations between each subscale are given in Table 6. The correlation coefficients are between .70 and .74. There is a positive correlation between each subscale. These results showed that this scale is reliable.

Table 6. Correlation coefficient of subscale

Subscale	*n*	*M*	*SD*	1	2	3
1. Confidence	1,332	8.97	2.49	1		
2. Persistence	1,332	9.37	2.43	.74**	1	
3. Flexible Thinking	1,332	10.08	2.34	.70**	.74**	1

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

Bebras Challenge scores from 30 students were used to evaluate the criterion-related validity of the scale. The Bebras Challenge test, based on informatics fundamentals, is a context for understanding students' computational thinking. To solve Bebras Challenge tasks, students need to demonstrate their ability to understand informatics fundamentals. They accomplish this by using information computation, data processing, data visualization, algorithm and programming concepts (Dagienė & Futschek, 2008). Our analysis showed a positive correlation between students' Bebras Challenge performance and their programming disposition scale ($r = .48$; $p < .01$). This result was in agreement with findings by Araujo et al. (2017) and arguing that the Bebras Challenge performance test was a good measure of students' aptitudes in computer science (Combéfis & Stupurienė, 2020). Therefore, programming dispositions correlates with computer science learning.

The construct validity of the scale shows that students' programming dispositions were accurately reflected and consistent with previous research findings and is consistent with respect to gender differences, academic track, and parental support (as cited in the previous sections). Gender differences with respect to programming dispositions are described as follows. Table 7 shows that male students ($M = 30.19$, $SD = 6.34$) had a higher programming disposition ($t = 8.32$; $p < .001$) than female students ($M = 27.22$, $SD = 6.45$). The result is consistent with the findings of previous studies which show that male students display more positive attitudes towards programming (Kong et al., 2018; Master et al., 2016). Male students, additionally, also displayed higher confidence, persistence and flexible thinking than did their female counterparts. This is consistent with previous research in computer science with respect to gender differences. Male students also had higher levels of confidence when encountering more difficult programming problems than female students (Settle et al., 2015). Katz et al. (2006) also showed that male students had a higher persistence in executing programming tasks than females. It is, consequently, important that these gender differences can be identified so that additional strategies can be developed to improve students' programming disposition: addressing the needs of both male and female students.

Table 7. Gender and programming disposition

Construct	Male		Female		*t*(744)	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Programming disposition	30.19	6.34	27.22	6.45	8.32	.000	0.33
Confidence	9.66	2.47	8.51	2.4	8.49	.000	0.33
Persistence	9.93	2.39	8.99	2.39	7.04	.000	0.28
Flexible thinking	10.61	2.21	9.72	2.35	6.91	.000	0.28

Another important variable to consider when examining the validity of this scale is that academic track of the individual student. Table 8 shows that students enrolled in a science track ($M = 30.26$, $SD = 6.42$) had a significantly higher programming disposition score ($t = 9.55$; $p < .001$) than students in a social science track ($M = 25.63$, $SD = 6.69$). In Taiwan, high school students in grades 11 and 12 are divided into two academic tracks: science and social science. High school students in the science track often enroll in additional science and advanced math courses in grades 11 and 12. Students in the social science track, however, tend to enroll in more social studies, humanity, and intermediate math courses rather than additional science and math courses. In this study, we found students with science background had more positive programming dispositions in all three constructs: confidence, persistence, and flexible thinking. The findings support the idea that the learning of programming is strongly linked with mathematical skills and abilities (Burrus & Moore, 2016) and science subjects (Durak & Saritepeci, 2018).

With regard to the role of parental support, our findings (r = .35, p < .01) are consistent with previous studies that showed a positive correlation between the degree of parental support and programming dispositions. Previous studies have shown that the more the parents valued programming activities, the more positive were the students' attitudes (Mason & Rich, 2020; Master et al., 2017). In this study, we investigated the link between parental support and students' programming dispositions. Our findings reveal that parental support shows a very definite positive correlation with programming dispositions. These findings are consistent with the results in the 2018 Programme for International Student Achievement and the 2019 findings of the Organization for Economic Cooperation and Development (OECD, 2019). The more support students got from parents, the higher the dispositions.

Table 8. Academic track and programming disposition

Construct	Science		Social science		*t*(1330)	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Programming disposition	30.26	6.42	25.63	6.69	9.55	.000	.50
Confidence	9.59	2.50	8.06	2.47	8.31	.000	.44
Persistence	10.00	2.38	8.28	2.54	9.40	.000	.49
Flexible thinking	10.68	2.23	9.28	2.45	8.05	.000	.42

Table 9 shows that students' programming dispositions were positively correlated with CAP mathematics scores. The CAP Chinese scores were, however, shown to be consistently negative. This shows that the programming disposition scale has a high discriminant validity. Erümit (2020) has indicated that mathematical activities had a positive effect on thinking flexibly for solving programming problems and persistence of programming learning. Katz et al. (2006) also found that learning experience of relevant subjects affected students' persistence of programming learning. In fact, the fields of programming and mathematics involve similar cognitive processes, such as logical thinking, computing, reasoning and problem solving. Therefore, mathematical ability and learning experiences are correlated with students' confidence in learning programming, persistence in facing complex tasks and thinking flexibly.

Table 9. Discriminant validity (N = 1,332)

	Programming disposition	Confidence	Persistence	Flexible thinking
Mathematics	$.12^{**}$	$.11^{**}$	$.11^{**}$	$.12^{**}$
Chinese	$-.12^{**}$	$-.13^{**}$	$-.11^{**}$	$-.08^{*}$

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

4.3. Confirmatory factor analysis (Q2)

CFA was used to examine the values of the goodness of fit. The results found χ^2 = 201.04 (df = 24), p < .001 (Figure 1). The results of the CFA did not show a good statistical fit probably due to our large sample size (over 200). For this reason, other statistical analyses needed to be used (Rigdon, 1995). The measurement for the goodness of fit here is composed of absolute fit indexes, relative fit indexes, and parsimonious fit indexes (Bagozzi & Yi, 1988). The analysis of this scale is: Absolute fit index: GFI = .97, AGFI = .94, RMR = .03, RMSEA = .07; relative fit index: NFI = .99, RFI = .98, CFI = .90; parsimonious fit indexes: PGFI = .52, PNFI = .66, CN = 312.84. This model passed all 10 standards (Table 10). In addition, the factor loading of all items were higher than the acceptable level (ranged from 0.57 to 0.86) (Figure 1). Item C2 (I can solve difficult programming tasks), P1 (When presented with a difficult programming task, I increase my efforts), FT1 (7. I would try alternative solutions when solving problems similar to the previous one) had the highest factor load in confidence, persistence and flexible thinking respectively. Table 11 shows the composite reliability [CR] > .7, average variance extracted [AVE] > 0.5. These results revealed this model was confirmed and produces high reliability and validity. The CFA supported that the construct of programming disposition is composed of confidence, persistence, and flexible thinking. "Confidence" among three constructs has the highest CR and AVE.

Figure 1. CFA diagram of the scale

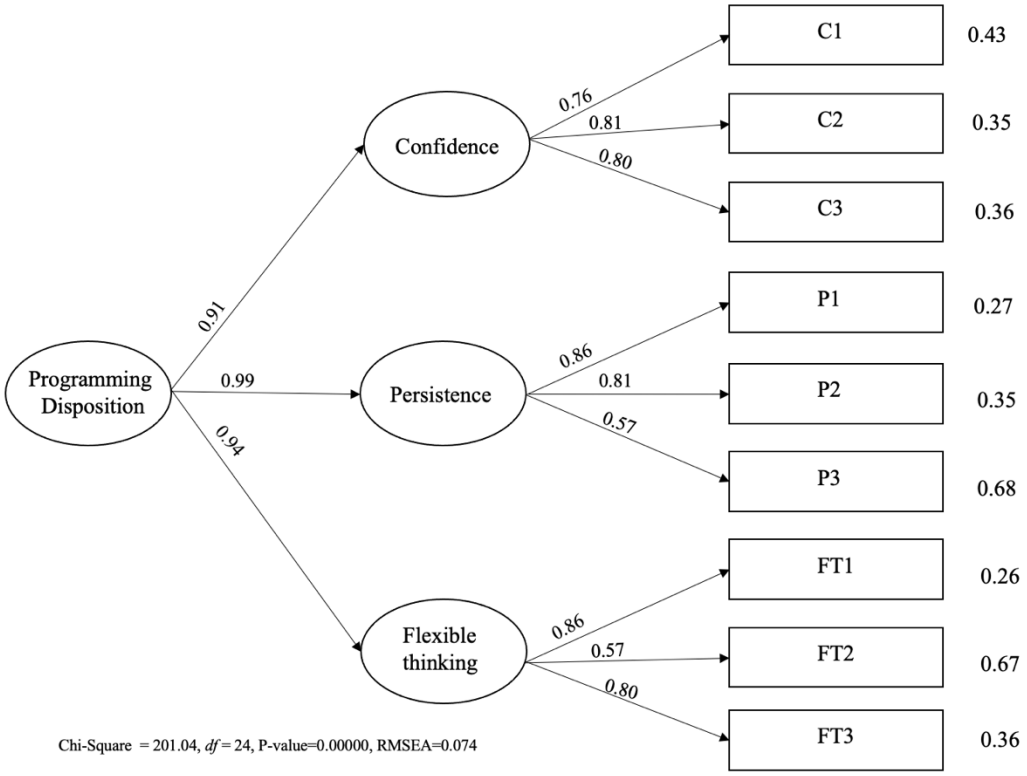


Table 10. Values of goodness of fit index

Goodness of fit index	Second-order factor analysis	Fit criteria of goodness of fit
Absolute fit index		
$\chi 2$	201.04	-
Df	24	-
GFI	0.97	> 0.9, good fit value
AGFI	0.94	> 0.9, good fit value
RMR	0.03	< 0.05, good fit value
RMSEA	0.07	< 0.08, reasonable fit value
Relative fit index		
NFI	0.99	> 0.9, good fit value
RFI	0.98	> 0.9, good fit value
CFI	0.99	> 0.9, good fit value
Parsimonious fit indexes		
PGFI	0.52	> 0.5
PNFI	0.66	> 0.5
CN	312.84	> 200 good sample quantity

Table 11. CR and AVE of three construct

Construct	CR	AVE
Confidence	0.83	0.62
Persistence	0.79	0.57
Flexible thinking	0.79	0.57

5. Significance of the programming disposition scale in educational settings

The significance of programming instruction has been addressed in literature (Burrus & Moore, 2016; Winslow, 1996). Previous findings have indicated that many students struggle with computer programming, which affects their engagement and motivation (Chookaew et al., 2015; Eliasson et al., 2006). Programming has a different nature from other disciplines because it involves both syntactic details and complex problem solving processes,

which requires intensive flexible thinking and persistence. Programming dispositions not only describes how much students are confident in programming, but also how they confront complicated problems. It also prescribes students' temperament of their roles when engaged in task performance (Association for Computing Machinery & IEEE Computer Machinery, 2020). Although previous research has devoted to studying effective instructional strategies for programming, it still lacks deeper exploration about students' disposition. Our research contributes to reveal more about students' behaviors and attitudes that characterize the inclination to carry out programming tasks.

The proposed programming disposition scale is an instrument for exploring how students communicate with programming tasks and their willingness to reflect on their own thinking and problem solving during programming. In educational settings, the programming disposition scale could be an effective tool for teachers to understand students' learning and evaluate the effects of instruction. Since the disposition scale moderates the behavior of applying knowledge and skills that becomes the context where and why the knowledge and skills are applied (Kusmaryono et al., 2019). This can, thus, be used as a guide for teachers to develop adaptive instruction to inspire and motivate their students for future studies or careers. A more inclusive learning environment can also be developed for students with different genders or from different cultures. In addition, teachers can evaluate whether their instructional strategies would inspire students' programming disposition, e.g., they can improve the implementation of STEM (Science, Technology, Engineering, Mathematics) education by considering programming disposition to arouse students' awareness of the integration of computational thinking and STEM disciplines.

6. Educational implications and suggestions

On the basis of research results, we produced some recommendations as follows.

Adaptive instruction: Among the three constructs, students' confidence was the lowest. This finding was similar to Mathematics. TIMSS (2020) released "TIMSS 2019 International Results in Mathematics and Science." The report showed that students in Taiwan often lack confidence in Mathematics and Science. Prior studies have proved that improper instructional design might lead to negative disposition (Katz, 1993). Therefore, instead of lecture-based instruction, more adaptive instruction should be provided based on students' characteristics. Exploration and experiment activities are effective for enhancing K-12 students' persistence and confidence. Through the process of struggling with complex problems, students' problem solving abilities can also be improved. Regarding gender issues, more adaptive learning activities should be designed to target to arouse females' interests and dispositions in programming. For example, Dagienė et al. (2015) found that through the task of dance moves, female students could understand better about the instructions or algorithm steps. Proper programming tools, such as visual programming, is also effective for engaging more females in programming (Baytak & Land, 2011; Kelleher & Pausch, 2006).

STEM instruction: The analysis of academic track showed the learning experience and ability of science and mathematics had a correlation with programming disposition. The knowledge and skills of science and mathematics should be integrated with programming practices. Lin et al. (2021) suggested that STEM education from multidisciplinary, such as programming and science, would increase students' interests. Erümit's (2020) study also pointed integrating mathematical activities into programming learning practices had a positive effect on thinking flexibly for solving programming problems and persistence of programming learning. Lin et al. (2019) found that through the STEM instruction, students had a higher confidence on programming learning. The results reveal "STEM" is an effective instructional strategy.

Jigsaw cooperative learning: Jigsaw cooperation is also an effective strategy for programming instruction. Teachers systemically divide learning tasks into different sub-tasks and assign students into groups, and each of the groups should complete one of the sub-tasks. Existing research has proved the effectiveness of Jigsaw strategies on students' knowledge building and confidence in programming (Garcia, 2021).

Parental support: Parental support has also been shown to be a vital factor in helping develop a student's programming disposition. It is imperative that schools help parents in understanding the importance of learning programming, a vital need for students' career development. Parents would not learn programming knowledge and skills but need to understand impacts of computing on daily life. The K-12 computer science framework listed three dimensions of impacts of computing: culture, social interaction and safety, law, and ethics. Thus, schools should conduct activities for parents to demonstrate the effects of computing, such as new cultural practices, equity and access to computing (K-12 Computer Science Framework Steering Committee, 2016).

7. Conclusion

This study develops a programming disposition scale for high school students. The scale is a five-point Likert-type scale and consists of 9 items. The internal consistency of the scale is excellent and the test-retest reliability is high. This scale appears to be respectably stable over time. The correlation coefficient of each subscale is positive. For the criterion validity, the scale shows a positive correlation with the Bebras Challenge. This scale also establishes the discriminant validity relevant to students' performance on mathematics and Chinese. The construct validity is validated by testing the variables of gender, academic track, and support from parents. The scale model has been verified by the CFA results. The structural equation modelling supports that the construct of programming disposition is composed of confidence, persistence, and flexible thinking. The results of these statistical analysis show that the scale is a valid and reliable tool. This study helps to expand our knowledge with respect to programming disposition, and improves the quality of assessment in programming education.

Our programming disposition scale has some strengths, such as filling an important gap in the field of assessment development for computer science education; teachers can utilize this scale to assess students' programming dispositions and find ways to help students learn in a programming course; students may, additionally, refer to the results on this scale and glean insights as to whether they should enroll in programming courses in high school or whether to choose related majors in a university setting; while this scale is primarily developed for students having experience in programming, it may also be useful for students who have little or no programming experience. However, a selection bias in participant recruitment might pose a threat to the internal validity of the research. All participants are in the Taipei metropolitan area in Taiwan. More subjects must be included in the future to get more generalized results for extending to other populations.

Acknowledgement

This research was sponsored by the Ministry of Science and Technology, Taiwan under Grant no. MOST 104-2511-S-003-055-MY3. We thank Dr. Ching-Lin Wu (Associate Professor at National Taiwan Normal University) for his assistance on this study.

References

Association for Computing Machinery & IEEE Computer Machinery. (2020). *Computing Curricula 2020: Paradigms for Global Computing Education*. https://www.acm.org/binaries/content/assets/education/curricula-recommendations/cc2020.pdf

Anastasiadou, S. D., & Karakos, A. S. (2011). The Beliefs of electrical and computer engineering students' regarding computer programming. *The International Journal of Technology, Knowledge and Society, 7*(1), 37-51. https://doi.org/10.18848/1832-3669/CGP/v07i01/56170

Araujo, A. L. S. O., Santos, J. S., Andrade, W. L., Guerrero, D. D. S., & Dagienė, V. (2017). Exploring computational thinking assessment in introductory programming courses. *Proceedings of 2017 IEEE Frontiers in Education Conference* (pp. 1-9). IEEE. https://doi.org/10.1109/FIE.2017.8190652

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74-94.

Barak, M., & Levenberg, A. (2016). Flexible thinking in learning: An Individual differences measure for learning in technology-enhanced environments. *Computers & Education, 99*, 39-52. https://doi.org/10.1016/j.compedu.2016.04.003

Baser, M. (2013). Attitude, gender and achievement in computer programming. *MiddleEast Journal of Scientific Research, 14*(2), 248–255. https://doi.org/10.5829/idosi.mejsr.2013.14.2.2007

Baytak, A., & Land, S. M. (2011). Advancing elementary-school girls' programming through game design. *International Journal of Gender, Science and Technology, 3*(1), 243–253.

Begel, A., & Nagappan, N. (2008). Pair programming: what's in it for me? *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement* (pp. 120-128). ACM. https://doi.org/10.1145/1414004.1414026

Bennedsen, J., & Caspersen, M. E. (2006). Abstraction ability as an indicator of success for learning object-oriented programming? *ACM SIGCSE Bulletin, 38*(2), 39-43. https://doi.org/10.1145/1138403.1138430

Biber, A. C., Tuna, A., & Incikabi, L. (2013). An Investigation of critical thinking dispositions of mathematics teacher candidates. *Educational Research, 4*(2), 109-117.

Breen, S., Cleary, J., & O'Shea, A. (2010). Measuring students' persistence on unfamiliar mathematical tasks. *Proceedings of the British Society for Research into Learning Mathematics, 30*(3), 19-24.

Burrus, J., & Moore, R. (2016). The Incremental validity of beliefs and attitudes for predicting mathematics achievement. *Learning and Individual Differences, 50*, 246-251. https://doi.org/10.1016/j.lindif.2016.08.019

Byrne, P., & Lyons, G. (2001). The Effect of student attributes on success in programming. *ACM SIGCSE Bulletin, 33*(3), 49-52. https://doi.org/10.1145/377435.377467

Charlton, J. P., & Birkett, P. E. (1999). An Integrative model of factors related to computing course performance. *Journal of Educational Computing Research, 20*(3), 237-257. https://doi.org/10.2190/BTG0-7VQK-6XD3-G4C4

Cheah, C. S. (2020). Factors contributing to the difficulties in teaching and learning of computer programming: A Literature review. *Contemporary Educational Technology, 12*(2), Article ep272. https://doi.org/10.30935/cedtech/8247

Chookaew, S., Wanichsan, D., Hwang, G. J., & Panjaburee, P. (2015). Effects of a personalised ubiquitous learning support system on university students' learning performance and attitudes in computer-programming courses. *International Journal of Mobile Learning and Organisation, 9*(3), 240-257. https://doi.org/10.1504/IJMLO.2015.074207

Combéfis, S., & Stupurienė, G. (2020). Bebras based activities for computer science education: Review and perspectives. In K. Kori & M. Laanpere (Eds.), *Lecture Notes in Computer Science: Vol. 12518. Informatics in Schools. Engaging Learners in Computational Thinking* (pp. 15–29). Springer. https://doi.org/10.1007/978-3-030-63212-0_2

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-type scales in the social sciences. *Journal of Adult Education, 40*(1), 19-22.

Dagienė, V., & Futschek, G. (2008). Bebras international contest on informatics and computer literacy: Criteria for good tasks. In R.T. Mittermeir, M.M. Sysło (Eds.), *Lecture Notes in Computer Science: Vol. 5090. Informatics Education - Supporting Computational Thinking* (pp. 15–29). Springer. https://doi.org/10.1007/978-3-540-69924-8_2

Dagienė, V., & Stupuriene, G. (2016). Bebras--A Sustainable community building model for the concept based learning of informatics and computational thinking. *Informatics in Education, 15*(1), 25-44. https://doi.org/10.15388/infedu.2016.02

Dagienė, V., Pelikis, E., & Stupurienė, G. (2015). Introducing computational thinking through a contest on informatics: Problem-solving and gender issues. *Information Sciences, 73*, 55-63.

Durak, H. Y., & Saritepeci, M. (2018). Analysis of the relation between computational thinking skills and various variables with the structural equation model. *Computers & Education, 116*, 191-202. https://doi.org/10.1016/j.compedu.2017.09.004

Eliasson, J., Westin, L. K., & Nordstrom, M. (2006). Investigating students' confidence in programming and problem solving. *Proceedings of 36th Annual Frontiers in Education Conference* (pp. 22-27). IEEE. https://doi.org/10.1109/FIE.2006.322490

Erümit, A. K. (2020). Effects of different teaching approaches on programming skills. *Education and Information Technologies, 25*(2), 1013-1037. https://doi.org/10.1007/s10639-019-10010-8

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for research in Mathematics Education, 7*(5), 324-326. https://doi.org/10.2307/748467

Fitzgerald, S., Lewandowski, G., McCauley, R., Murphy, L., Simon, B., Thomas, L., & Zander, C. (2008). Debugging: Finding, fixing and flailing, a multi-institutional study of novice debuggers. *Computer Science Education, 18*(2), 93-116. https://doi.org/10.1080/08993400802114508

Galgouranas, S., & Xinogalos, S. (2018). jAVANT-GARDE: A Cross-platform serious game for an introduction to programming with Java. *Simulation & Gaming, 49*(6), 751-767. https://doi.org/10.1177/1046878118789976

Garcia, M. B. (2021). Cooperative learning in computer programming: A Quasi-experimental evaluation of Jigsaw teaching strategy with novice programmers. *Education and Information Technologies, 26*, 1-18. https://doi.org/10.1007/s10639-021-10502-6

Golding, P., Facey-Shaw, L., & Tennant, V. (2006). Effects of peer tutoring, attitude and personality on academic performance of first year introductory programming students. *Proceedings of 36th Annual Frontiers in Education Conference* (pp. 7-12). IEEE. https://doi.org/10.1109/FIE.2006.322662

Gomes, A. J., Santos, A. N., & Mendes, A. J. (2012). A Study on students' behaviours and attitudes towards learning to program. *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education* (pp. 132-137). ACM. https://doi.org/10.1145/2325296.2325331

Grandell, L., Peltomäki, M., & Salakoski, T. (2005). High school programming—A Beyond-syntax analysis of novice programmers' difficulties. *Proceedings of the Koli Calling 2005 Conference on Computer Science Education* (pp. 17-24). ACM.

Gray, W. D., Goldberg, N. C., & Byrnes, S. A. (1993). Novices and programming: Merely a difficult subject (why?) or a means to mastering metacognitive skills? A Review of Soloway's & Spohrer's, Studying the Novice Programmer. *Journal of Educational Research on Computers*, *9*(1), 131-140.

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A Review of the state of the field. *Educational Researcher, 42*(1), 38-43. https://doi.org/10.3102/0013189X12463051

Hu, C. F., Wang, A. T., Wu, C. C., & Lin, Y. T. (2020). Identifying learning concepts for the new 12-year basic education ICT curriculum: A Delphi survey. *Bulletin of Educational Research, 66*(1), 77-102. https://doi.org/10.3966/102887082020036601003

Hu, C. F., Wu, C.C., Lin, Y. T., & Yu, C. H. (2021). Development of a computational thinking test for high school students. *International Journal on Digital Learning Technology, 13*(1), 1-21. https://doi.org/10.3966/2071260X2021011301001

International Society for Technology in Education & the Computer Science Teachers Association (ISTE & CSTA). (2011). *Operational Definition of Computational Thinking for K–12 Education.* https://cdn.iste.org/www-root/Computational_Thinking_Operational_Definition_ISTE.pdf

Jang, I. O., & Lew, H. C. (2014). Case studies in thinking processes of mathematically gifted elementary students through Logo programming. *Work, 4*, 9.

Jiau, H. C., Chen, J. C., & Ssu, K.-F. (2009). Enhancing self-motivation in learning programming using game-based simulation and metrics. *IEEE Transactions on Education, 52*(4), 555-562. https://doi.org/10.1109/TE.2008.2010983

K-12 Computer Science Framework Steering Committee. (2016). *K-12 Computer science framework.* https://k12cs.org/wp-content/uploads/2016/09/K-12-Computer-Science-Framework.pdf

Katz, L. G. (1993). *Dispositions as educational goals* (ED363454). ERIC. https://files.eric.ed.gov/fulltext/ED363454.pdf

Katz, S., Allbritton, D., Aronis, J., Wilson, C., & Soffa, M. L. (2006). Gender, achievement, and persistence in an undergraduate computer science program. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 37*(4), 42-57. https://doi.org/10.1145/1185335.1185344

Kelleher, C., & Pausch, R. (2006). Lessons learned from designing a programming system to support middle school girls creating animated stories. In J. Grundy & J. Howse (Eds.), *Proceedings of Visual Languages and Human-Centric Computing* (pp. 165–172). IEEE. https://doi.org/10.1109/VLHCC.2006.30

Kong, S. C., Chiu, M. M., & Lai, M. (2018). A Study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education. *Computers & Education, 127*, 178-189. https://doi.org/10.1016/j.compedu.2018.08.026

Kusmaryono, I., Suyitno, H., Dwijanto, D., & Dwidayati, N. (2019). The Effect of mathematical disposition on mathematical power formation: Review of dispositional mental functions. *International Journal of Instruction, 12*(1), 343-356.

Lin, Y. T., Wang, M. T., & Wu, C. C. (2019). Design and implementation of interdisciplinary STEM instruction: Teaching programming by computational physics. *The Asia-Pacific Education Researcher, 28*(1), 77-91. https://doi.org/10.1007/s40299-018-0415-0

Lin, Y. T., Yeh, M. K. C., & Hsieh, H. L. (2021). Teaching computer programming to science majors by modelling. *Computer Applications in Engineering Education, 29*(1), 130-144. https://doi.org/10.1002/cae.22247

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12?. *Computers in Human Behavior, 41*, 51-61. https://doi.org/10.1016/j.chb.2014.09.012

Luxton-Reilly, A. (2016). Learning to program is easy. *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 284-289). ACM. https://doi.org/10.1145/2899415.2899432

Margulieux, L., Ketenci, T. A., & Decker, A. (2019). Review of measurements used in computing education research and suggestions for increasing standardization. *Computer Science Education, 29*(1), 49-78. https://doi.org/10.1080/08993408.2018.1562145

Mason, S. L., & Rich, P. J. (2020). Development and analysis of the Elementary Student Coding Attitudes Survey. *Computers & Education, 153,* Article 103898. https://doi.org/10.1016/j.compedu.2020.103898

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology, 108*(3), 424-437. https://doi.org/10.1037/edu0000061

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology, 160*, 92-106. https://doi.org/10.1016/j.jecp.2017.03.013

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. https://www.nctm.org/Standards-and-Positions/More-NCTM-Standards/Curriculum-and-Evaluation-Standards-(1989)

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. https://www.nctm.org/uploadedFiles/Standards_and_Positions/PSSM_ExecutiveSummary.pdf

Organization for Economic Cooperation and Development (OECD). (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. https://www.oecd.org/pisa/publications/pisa-2018-results-volume-ii-b5fd1b8f-en.htm

Pérez, A. (2018). A Framework for computational thinking dispositions in mathematics education. *Journal for Research in Mathematics Education, 49*(4), 424-461. https://doi.org/10.5951/jresematheduc.49.4.0424

Pioro, B. T. (2006). Introductory computer programming: Gender, major, discrete mathematics, and calculus. *Journal of Computing Sciences in Colleges, 21*(5), 123-129.

Rigdon, E. (1995). A Necessary and sufficient identification rule for structural equation models estimated. *Multivariate Behavioral Research, 30*, 359-383. https://doi.org/10.1207/s15327906mbr3003_4

Royster, D. C., Kim Harris, M., & Schoeps, N. (1999). Dispositions of college mathematics students. *International Journal of Mathematical Education in Science and Technology, 30*(3), 317-333. https://doi.org/10.1080/002073999287851

Rum, S. N. M., & Ismail, M. A. (2017). Metocognitive support accelerates computer assisted learning for novice programmers. *Educational Technology & Society, 20*(3), 170-181.

Sáez-López, J. M., Román-González, M., & Vázquez-Cano, E. (2016). Visual programming languages integrated across the curriculum in elementary school: A Two year case study using "Scratch" in five schools. *Computers & Education, 97*, 129-141. https://doi.org/10.1016/j.compedu.2016.03.003

Settle, A., Lalor, J., & Steinbach, T. (2015). Reconsidering the impact of CS1 on novice attitudes. In A. Decker (Eds.),*Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (pp. 229-234). ACM. https://doi.org/10.1145/2676723.2677235

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*(2), 342–357. https://doi.org/10.1037/0022-0663.89.2.342

Stokoe, R. (2012). Curiosity, a condition for learning. *The International Schools Journal, 32*(1), 63-65.

Sun, J. C. Y., & Hsu, K. Y. C. (2019). A Smart eye-tracking feedback scaffolding approach to improving students' learning self-efficacy and performance in a C programming course. *Computers in Human Behavior, 95*, 66-72. https://doi.org/10.1016/j.chb.2019.01.036

Tew, A. E., & Dorn, B. (2013). The Case for validated tools in computer science education research. *Computer, 46*(9), 60-66. https://doi.org/10.1109/MC.2013.259

Trends in International Mathematics and Science Study and the Progress in International Reading Literacy Study International Study Center (TIMSS). (2020). *TIMSS 2019 International Results in Mathematics and Science*. https://timssandpirls.bc.edu/timss2019/international-results/

Tsai, M. J., Wang, C. Y., & Hsu, P. F. (2019). Developing the computer programming self-efficacy scale for computer literacy education. *Journal of Educational Computing Research, 56*(8), 1345-1360. https://doi.org/10.1177/0735633117746747

Whitin, P. E.(2007). The Mathematics survey: A Tool for assessing attitudes and dispositions. *Teaching Children Mathematics, 13*(8), 426-433. https://doi.org/10.5951/TCM.13.8.0426

Wilkins, J. L. (2000). Preparing for the 21st century: The Status of quantitative literacy in the United States. *School Science and Mathematics, 100*(8), 405-418. https://doi.org/10.1111/j.1949-8594.2000.tb17329.

Winslow, L. E. (1996). Programming pedagogy—A Psychological overview. *ACM Sigcse Bulletin, 28*(3), 17-22. https://doi.org/10.1145/234867.234872

Yurdugül, H., & Aşkar, P. (2013). Learning programming, problem solving and gender: A Longitudinal study. *Procedia-Social and Behavioral Sciences, 83*, 605-610. https://doi.org/10.1016/j.sbspro.2013.06.115

How Roles in Collaboration Respond to the Exchange of Device-Student Ratio Under the Impact of External Scripts?

Wenhui Liu and Cixiao Wang*
School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing, China // 202031010056@mail.bnu.edu.cn // wangcixiao@bnu.edu.cn
*Corresponding author

(Submitted April 28, 2021; Revised September 9, 2021; Accepted September 13, 2021)

ABSTRACT: In science classrooms, technology affordance varies depending on device-student ratios (DSR) and the ways virtual manipulatives on mobile devices are used. Additionally, external scripts (ES) are widely used to promote effective group interaction in collaborative learning. Therefore, this research explored the influence of DSR and ES on collaborative inquiry learning. This research adopted a counterbalanced design between two rounds of experiments. A total of 128 students (including 11 dropouts) from four sixth-grade classes participated, with the four classes randomly divided into four experimental groups. Thematic analysis, social network analysis, and statistical analysis methods were used to analyze the distribution and transition of roles, the interaction between roles, and the self-efficacy and collective efficacy of the roles. The results illustrated that the role distribution was affected by DSR and ES, and frequent transitions of operational roles in groups emerged when DSR was exchanged. Moreover, the role of ES was reported in this study; it promoted the stability of role interaction on the one hand while significantly promoting self-efficacy and collective efficacy on the other. The study also proposed that the discourse statuses of different roles in collaborative learning were significantly different, and roles with a weaker discourse status had lower self-efficacy.

Keywords: Collaborative inquiry learning, Technology affordance, Role interaction, Self-efficacy, Social network analysis

1. Introduction

The term "technology affordances" refers to the interactive relationship between the actor and technology (Jeong & Hmelo-Silver, 2016). The affordance of learning technology has been widely discussed in the field of computer-supported collaborative learning (CSCL). When provided with different types of technological support, the learning process varies due to students' perceived affordance in terms of information acquisition, resource sharing, and process management (Jeong & Hmelo-Silver, 2016). Virtual manipulatives (VM) are computer-based simulations of physical manipulatives (PM) that can be accessed via the Internet or computer software (Bouck & Flanagan, 2009). In science courses, VM running on mobile devices make science inquiry learning flexible and convenient (Jou et al., 2016). Additionally, researchers recognized the role of the mobile device-student ratio on the learning process. Students have better learning experiences and performance with a 1:1 device-student ratio (DSR) due to seamless resource acquisition anytime and anywhere (e.g., Looi et al., 2011; Wong & Looi, 2011). However, some argue that different DSR provides different interaction and collaboration dynamics, for example, a group sharing one device (1:m) may facilitate a full discussion (Lin, et al., 2012). How the DSR influences the interaction patterns of groups remains rare in face-to-face science classrooms. To reveal that the group interaction varies under the impact of the DSR, we considered two conditions: 1:1 and 1:m.

In CSCL, scripts have been widely used to facilitate effective interactions. As socio-cognitive scaffolds for learners (Vogel et al., 2016), external scripts (ES) could structure sequences of activities and role distributions among group members (King, 2007). However, the adverse effects of scripting have been reported in some studies (e.g., Strijbos & Weinberger, 2010). For example, over-scripting may be thought to impede transfer or the development of self-regulatory abilities (Goodyear et al., 2014). Additionally, some studies have indicated ES could moderate the effect of DSR on the learning performance of collaborative inquiry learning (Wang & Le, 2021). The actual interactive process in collaborative learning activities with ES is not clear (Vogel et al., 2016). Therefore, in addition to the DSR, this study considers the effects of ES (with or without) on group interaction patterns in collaborative inquiry learning.

Social interactions in collaborative learning have been a hot topic in recent years. Previous studies focus more on learning outcomes when using VM in collaborative inquiry activities (e.g., Zacharia & Olympiou, 2011). Less attention has been given to how community members communicate with each other under various technical

environments. Additionally, roles (e.g., leader, recorder) played by group members could reflect the interaction patterns (e.g., role distribution, role interaction) during group work (Simpson et al., 2017; Kirschner et al., 2018). Moreover, self-efficacy and collective efficacy are important motivational factors that influence group interaction behavior (Wang & Lin, 2007), and they are also affected by the external environment provided. Thus, the purpose of this study is to reveal the effects of DSR and ES on role interactions, self-efficacy, and collective efficacy, as well as the correlation between them.

2. Literature review

2.1. Technology affordance for collaborative inquiry learning

Gibson (1977) originally defined affordances, referring to possibilities of action provided to an individual by the environment. In CSCL, affordances vary depending on the cultural-symbolic and material properties of the technology (Vyas et al., 2006), as well as how the affordances relate to learners' relevant experience (Jeong & Hmelo-Silver, 2016). Students thus develop perceived affordances for goal-oriented actions when given constraints in a dynamic learning environment (Abrahamson & Sánchez-García, 2016). However, affordances of learning technology need to be examined and understood in terms of what they enable learners to do for greater learning outcomes (Jeong & Hmelo-Silver, 2016).

Collaborative inquiry learning (CIL) originates from the actual activities of scientific inquiry in which students in groups engage in self-regulated learning activities supported by the teacher, and its key features include orientation and questioning, hypothesis generation, planning, investigation, analysis, and interpretation of data, model exploration and creation, conclusion and evaluation, communication, and prediction (Bell et al., 2010). In science classrooms, VM used on tablet PCs is a popular inquiry tool for CIL (Fokides & Mastrokoukou, 2018). Researchers have revealed that VM carries unique affordances compared to PM, including safe and cost-efficient (Hsu & Thomas, 2002), visualization and repeatability (Olympiou & Zacharia, 2012), and flexibility and convenience, especially with the widespread application of mobile technology (e.g., tablet PCs, interactive tabletops) (Jou et al., 2016).

Furthermore, the DSR is an important factor in designing the learning technology environment, especially if the screens are small. For example, Hassler et al. (2016) developed the view that tablets may be best suited for individual rather than collaborative use because of customizability, and some students may be reluctant to share tablets. Researchers have shown that 1:1 DSR could enhance students' learning experience (e.g., Looi et al., 2011; Wong et al., 2010). Lin et al. (2012) summarized the special affordances of the 1:1 DSR as follows: perpetual and ubiquitous learning, authentic and contextualized learning, seamless learning, and rapid knowledge co-construction. Nevertheless, Dillenbourg and Evans (2011) revealed that, while desktops (1:1) are personal, tabletops (1:m) could afford collaboration and facilitate the emergence of different viewpoints in student groups.

Some studies have contrasted students' performances in collaborative learning between the 1:1 and 1:m DSR. For example, Lin et al. (2012) performed quasi-experimental research investigating the effects of two settings on CCM (collaborative concept mapping) in two 6th-grade classes. The findings indicated that the 1:m groups generated concept maps superior to the 1:1 groups due to well-discussed notes, while the 1:1 groups demonstrated better quality interactions. However, the reason why quality interactions did not lead to good production remains unclear. The resource allocation and interaction process may be the factors that influence learning performances in terms of the DSR. However, there are few studies on that topic, and more details need to be revealed.

2.2. External scripts in collaborative learning

Students rarely spontaneously generate effective interaction in collaborative learning without some form of guidance (Dillenbourg & Hong, 2008; King, 2007; Kollar et al., 2006). In CSCL, scripts could facilitate the process of collaboration by structuring interactive processes, sequencing the activities, and guiding the discussion (Kollar et al., 2006; Vogel et al., 2016). The learning activities induced by scripts (e.g., roles of participants, the actions engaged in) could prompt specific cognitive, socio-cognitive, and metacognitive processes and closely relate to the intended learning (King, 2007).

There are two kinds of collaboration scripts: internal scripts and external scripts. External scripts (ES) are usually provided by teachers or learning facilitators to prompt group interaction (King, 2007) not otherwise represented

in learners' cognitive systems but rather in their external surroundings, typically at the beginning of a collaborative learning situation (Kollar et al., 2006). In contrast, internal scripts often refer to an internalized version of an ES. It may also refer to prior socially/culturally derived rules for cooperating (King, 2007). Thus, internal scripts exhibited interindividual differences with respect to their degree of structuredness, which was related to individuals' concrete experiences with situations (Kollar et al., 2006). Dillenbourg identified five types of ES according to the level of coercion. Specifically, induced scripts convey the designer's expectations implicitly; instructed scripts are oral or written instructions of the teacher's expectations by which students can construct an internal script that corresponds to the ES; trained scripts train students to collaborate in a specific way and may have control over the student's internal script; prompted scripts provide cues that direct students to specific roles; follow-me scripts strictly control how students interact with an environment (Dillenbourg & Jermann, 2007). Instructed and prompted scripts are often used in CIL of science courses, not only allow students to develop their internal scripts, but also facilitate group interaction by special roles.

Previous studies have found that CSCL with ES had a large positive effect on the collaboration skills and learning outcomes of students in domain-specific knowledge areas (Radkowitsch et al., 2020; Vogel et al., 2016). However, ES may influence naturally emerging collaboration and are not conducive to self-regulatory abilities (Goodyear et al., 2014). Therefore, this study will also explore the effect of ES on CIL.

2.3. The roles, self-efficacy, and collective efficacy

The assignment of group roles is an important indicator of the performance of group coordination (Strijbos et al., 2007), can promote the formation of positive interdependence, and contributes to the success of collaboration (Antle, 2014). There are two perspectives about roles in CSCL, namely, emerging roles and scripted roles. Emerging roles are spontaneously developed by participants, emphasize the learner's structure, and self-regulate their CSCL processes (Strijbos & Weinberger, 2010). The development of emerging roles is dynamic over longer periods according to learners' knowledge acquisition and collaborative learning experiences (Strijbos & Weinberger, 2010). However, spontaneous role participation is often unequal participation (Simpson et al., 2017; Lloyd & Cohen, 1999). For example, Simpson et al. (2017) detected uneven participation patterns, unequal status orderings, and an imbalance of power in group interactions. In contrast, script roles are assigned to learners by educational designers to promote equal participation of group members and structured collaboration. Both emerging roles and script roles could influence the interaction patterns of collaborative groups; thus, we will discuss their interactions.

Specifically, what roles will appear in the process of collaboration? In face-to-face CIL, each role has implicit responsibility for information processing, cognitive participation, and so on, but to complete a task together, each role has explicit behavior responsibility. Accordingly, role categorization of function may reflect group interaction patterns. Johnson and Johnson (1987) classified the functions of roles in group collaboration into four categories: formative, functional, summative, and promotive. An example of a formative role could be order supervisor, while functional roles include recorder, motivator, clarifier, interpreter, and consensus seeker. Summative roles contain summarizer and creator, while the duties of promotive roles include reason requesting and principle giving (Johnson & Johnson, 1987). Furthermore, Wang (2021) identified five functional role categories in CIL in science courses, including coordinator, integrator, inquirer, facilitator, and marginal. The categories will be used in this study to identify roles that emerge in CIL.

Self-efficacy (SE) and collective efficacy (CE) are the motivating factors for collaborative learning. SE (students' perceptions of their capability to achieve the desired outcome) is specifically related to the learning experience and the external environment (Bandura, 1997) and may also affect learning cognitive processes (Girasoli & Hannafin, 2008) and academic achievement (e.g., Tsai et al., 2011). In CSCL, researchers have suggested that collaborative learning behavior may be influenced by the SE of group members (Wang & Lin, 2007). CE refers to the group's shared beliefs in its conjoint capabilities to execute the courses of action required to achieve designated goals (Bandura, 1997). According to Wang et al. (2014) and Wang and Lin (2007), CE has a positive impact on group performance and could be affected by group interaction behaviors due to students with higher collective efficacy using more high-level cognitive skills in group discussions (Wang et al., 2014). However, few researchers have explored the influence of different technology affordances on SE and CE in collaborative learning. Therefore, in addition to examining the involvement and interaction of roles, this study investigates the SE and CE of roles in different technology affordances.

2.4. Purposes of this study

We conducted two-round CIL activities in a primary school to depict how roles variously respond to the DSR exchange under the impact of ES. There are two independent variables: device-student ratio (DSR) and external script (ES). The symbol 1:1 DSR represents each student has a tablet, while the 1:m DSR demonstrates only one tablet for each collaborative group. The mark (with or without) under ES refers to whether external support was provided to facilitate collaboration through structuring the interactive processes. In terms of dependent variables, we discussed four aspects: role distribution, role interaction, self-efficacy (SE), and collective efficacy (CE). The Roles refer to each group member's responsibilities, which include coordinator, integrator, inquirer, facilitator, and marginal. The proportion of five role categories that emerge during inquiry processes is referred to as role distribution. Role interaction refers to the closeness of interaction among group members and the interaction patterns of groups. SE and CE reflect learners' beliefs about individual and group capability in CIL. In summary, four main research questions are addressed as follows:

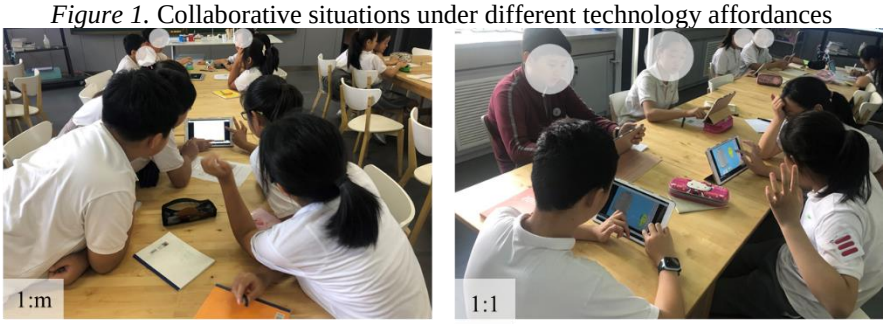
- How did the DSR exchange impact role distributions with and without ES?
- How did the DSR exchange impact role interactions with and without ES?
- How did the DSR exchange impact SE and CE with and without ES?
- What is the relationship between roles, SE, and CE?

3. Experimental design

3.1. Participants

The study involved 128 children, aged 10-12, in sixth grade (69 girls, 59 boys). They were from four classes of a public primary school located in Beijing, China, and taught by the same science teacher. An informed consent form that included detailed information about that experiment was signed by all participants, parents, and science teachers. We termed these four classes Class A (15 boys, 16 girls), Class B (16 boys, 18 girls), Class C (16 boys, 16 girls), and Class D (12 boys, 19 girls). Figure 1 depicts the collaborative situations in the two DSR conditions. Students in each class were randomly assigned to 6 learning groups, with 5-6 students in each group. These groups were long-term cooperative learning groups and had loyal interpersonal relations and adequate prior team experience (Johnson & Johnson, 1999). Additionally, there was no significant difference in the science exam in the last academic year among these four classes (p = .167 > .05). It should be noted that 11 students did not accomplish all experimental procedures due to various factors, such as time conflicts, students' physical conditions, and learning situations. Thus, the error caused by the 11 dropout students could be viewed as a limitation in this study.

Figure 1. Collaborative situations under different technology affordances



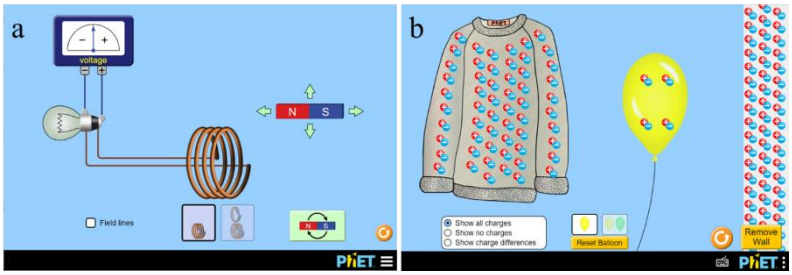
3.2. Inquiry technologies

3.2.1. Virtual manipulatives

The scientific inquiry theme in the first round is electromagnetic induction (Mag.), and the theme after affordance exchange is triboelectrification (Fri.). They were chosen from the contents of the fifth- and sixth-grade science curriculum according to the syllabus of the selected school. Accordingly, VM was selected from the PhET learning platform (phet.colorado.edu), a free online simulation program. These two topics are related and have the same level of difficulty, and the operation of VM is essentially the same. These simulations run independently and separately on tablets running on Android, with a screen size of 8 inches and a screen ratio of 16:10 (see Figure 2). Before conducting experiments, all participating students had prior experience using these

tablets for learning. In 1:1 DSR, each student could operate and observe using his or her own tablet anytime and anywhere. However, group members in 1:m DSR share one tablet for operation and observation, necessitating negotiation over the use of restricted resources.

Figure 2. Interactive interface of VM: a. Mag; b. Fri.



3.2.2. External scripts

This study designed two types of ES. Instructed scripts were guidelines for the CIL provided by the science teacher through PowerPoint to students (see Figure 3), while prompt scripts were cues that inspired students to play a particular role (see Figure 4). During the collaboration, students could play most of the roles on their own according to the prompt. However, for all group members' discussions, the scripted role of "inspector" was delegated to particular students.

Figure 3. Instructed script

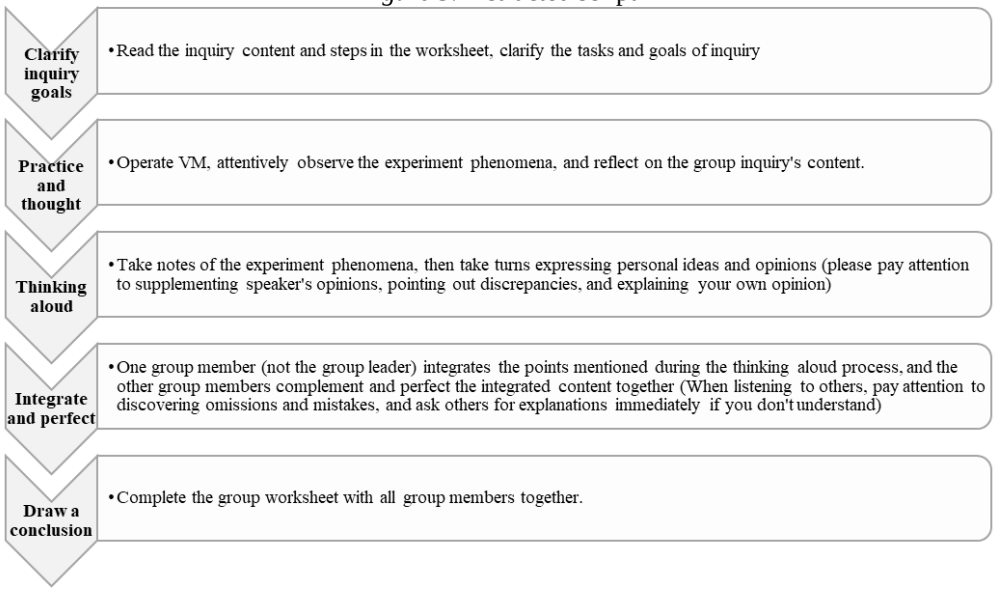


Figure 4. Prompt script

The function of Inspector

- Keep the discussion from getting sidetracked
- Remind group members to follow the experiment requirements
- Manage the inquiry process and allocate time reasonably
- Facilitate all group members to participate in the collaborative inquiry

3.2.3. Group worksheet

Students were also driven by group worksheets jointly developed by the science teacher and the researchers in conformity with the science curriculum standards. In each group worksheet, there were two inquiry tasks. They

were designed to examine the students' concept interpretation and problem-solving competence, respectively. The worksheet's total score was 100 points, with 50 points each for the two tasks. We also obtained the group members' seat details from the worksheets.

3.3. Procedure

In keeping with the school's science curriculum, this research was performed in two-round in June and September 2019. In the first round, the inquiry situations of the four classes were as follows: A: 1:1 DSR and with ES; B: 1:m DSR and with ES; C: 1:1 DSR and without ES; and D: 1:m DSR and without ES (see Figure 5). In the second round, the status of DSR for each class was exchanged, such as class A/C from 1:1 to 1:m and class B/D from 1:m to 1:1; simultaneously, the availability of ES remained the same.

Figure 5. Diagram of counterbalanced design

	Class A	Class B	Class C	Class D
Mag.	1:1 DSR with ES	1:m DSR with ES	1:1 DSR without ES	1:m DSR without ES
⇩				
Fri.	1:m DSR with ES	1:1 DSR with ES	1:m DSR without ES	1:1 DSR without ES

Inquiry learning activities have three steps. The teacher introduces the inquiry task in the first step (10min). Specifically, the teacher presents and explains the instructed script and distributes the prompt script to groups with ES. Step two is for groups to do independent inquiry and finish the worksheet without the teacher's intervention (25min). The final step is to instruct all students to complete a reflection questionnaire (10min).

3.4. Instruments

A reflection questionnaire, created by two primary science teachers and one primary Chinese teacher, was used to collect the relevant information of students during the inquiry processes. This questionnaire consists of two parts: (1) three open-ended questions about roles and (2) two efficacy items that used a seven-point Likert scale. The three open-ended questions are as follows: (a) What tasks and roles did you play in the group inquiry? (b) How did your group use the tablet(s) to collaborate? (c) Which group members did you communicate more with? Three questions were used to investigate the emerged roles, the way students used learning resources, and the interaction among group participants during the group inquiry process.

The items of SE and CE were inspired by the PASS scale established by Pass in 1992. This scale is probably the most commonly used subjective appraisal scale. It contains only a nine-point scale item, namely, "How much mental effort did you put into the learning process?" (Paas, 1992). Thus, this study used two seven-point scale items: (a) "My success in this task in group collaboration was…"; and (b) In this task, my group's performance in the class was…". The options ranged from (1) "very poor" to (7) "very good".

3.5. Data analysis

3.5.1. Thematic analysis

A qualitative thematic approach was applied to analyzing role information reported by students in the reflection questionnaire. The role information was encoded based on the codes proposed by Wang (2021). In this study, first, the work content with similar characteristics was summarized into abstract role labels; second, the role labels were classified into five categories (see Table 1). We tested the consistency of role coding results of the two researchers for the two-round experiments. The inter-rater reliability Kappa was .839 ($p < .001$, $N = 245$), which indicates good consistency. The science teacher was invited to examine the divergence between two researchers' coding results and collaborate with the researchers to unify the inconsistency.

Table 1. The role labels and categories in this study

Role category	Role label	Role definition	Examples of students description
Coordinator (R1)	Leader	The group leader is assigned by the teacher and leads the group to complete the task	"lead group members to explore learning tasks"
	Inspector	Script role (Figure 4)	"identify the content of exploration and prevent discussion deviation"
Integrator (R2)	Recorder	Recording inquiry processes and findings according to the group worksheet	"participate in discussions and record answers"
	Integrator	Summarizing ideas and suggestions	"summarize"
	Leader & Recorder	Acts as both leader and recorder	"the group leader, and wrote worksheet"
	Inspector & Recorder	Acts as both inspector and recorder	"I was an inspector and wrote worksheets meantime"
Inquirer (R3)	Investigator	Proposing solutions to problems and drawing conclusions	"explore to find the answer to the question"
	Experimenter	Operating the tablet(s) to conduct the experiment	"primary operator, operating tablet computers"
Facilitator (R4)	Proposer	Offering suggestions	"I'm an idea person to the group"
	Facilitator	Assisting group members in completing the inquiry tasks	"assist others"
Marginal (R5)	Observer	Just observing the whole process of inquiry	"listen to the team members"
	Spectator	Regarding oneself as the spectator/outsider of the group	"audience, watching other people" "play soy sauce", "silent bear"
	(null)	Students did not report the played role	

3.5.2. Social network analysis

In this study, the social network analysis method was applied to analyze the interaction of group members' roles. The data source is the actual seating information filled in by the students in group worksheets, that is, the proximity relationship of group members, as well as the member communication reported in the reflection questionnaire. We use two concepts (density and component) to extract the social network characteristics of groups. First, network density represents the closeness of interactions among group members. The value range of density is [0, 1], and the more connections between nodes, the higher the network density is. The network density of the social network graph with directional arrows can be expressed by the following formula, where the symbol n represents the number of nodes and the symbol d_i represents the degree of node i.

$$\sum d_i / n(n-1)$$

Second, the "action-oriented component" was used to analyze the focus of the group work in the inquiry activity. A component refers to the subgraph of a social network graph with all nodes connected (Marin & Wellman, 2011). This study selected the in-degree of nodes to extract the components of the directed graph. To find the "action-oriented component," initially, the node with the highest in-centrality in a group was marked. Then, starting from this node, we found the component with 3 nodes and selected the component with the highest in-centrality degree. If multiple components with the highest in-centrality degree were found, the component with the most two-way connections was selected. If a group did not contain such components, this group could be considered a non-oriented group. Based on identified network components, this study divided the action orientation of participating groups into three types: task-oriented, inquiry-oriented, and non-oriented. For visualization, they were assigned values of 1, -1, and 0.

- Task-oriented group: the action-oriented component contained R2, and the in-degree of R2 was not 0. These groups focused on completing group worksheets.
- Inquiry-oriented group: the action-oriented component contained R3 but did not contain R2, or the in-degree of R2 was 0. These groups focused on operating the VM and conducting an inquiry.

- Non-oriented group: the action-oriented component contains R2 roles with a 0 in-degree of recorders but no R3 roles; or does not contain R2 and R3 roles; or there was no action-oriented component.

3.5.3. Statistical analysis

As the SE and CE scores of students do not meet normal distribution, respectively nonparametric tests such as the Jonckheere-Terpstra test, the Mann-Whitney U test, and the Wilcoxon signed-rank test were used for revealing differences. Additionally, the group learning performance (the score of the group worksheet) was analyzed using two-way ANOVA. The statistical analysis was performed using SPSS24.0 with $p < .05$ defined as statistically significant.

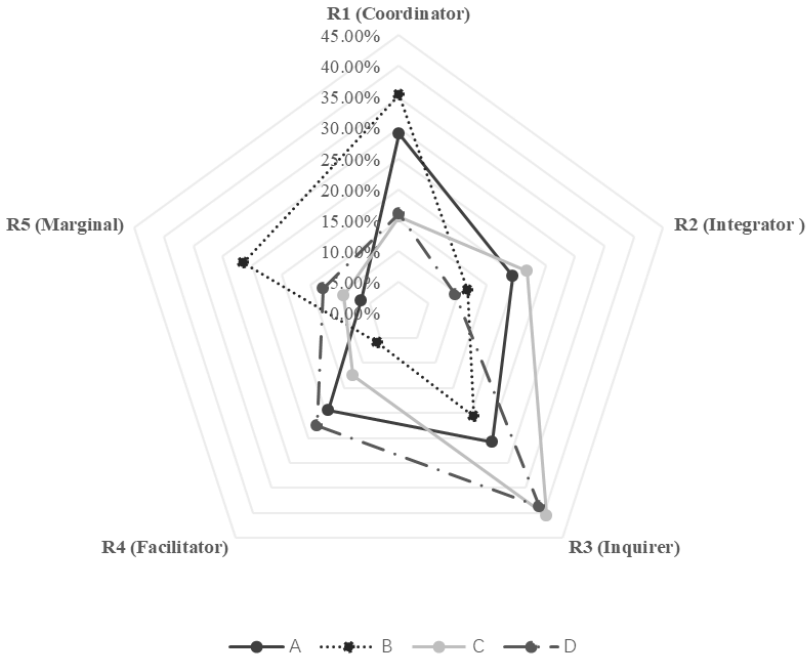
4. Results

4.1. Role distribution and transition

4.1.1. Role distribution

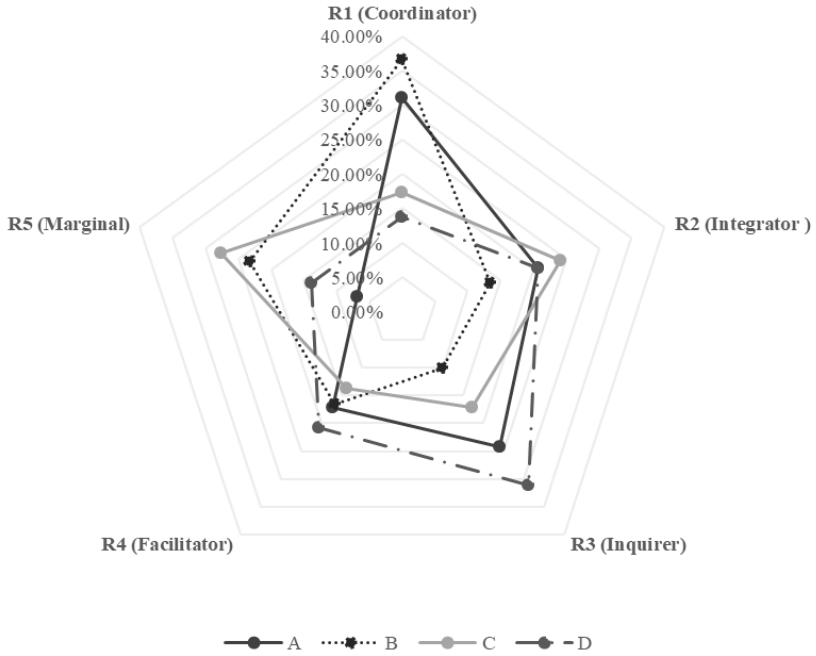
According to the results of thematic analysis, the proportion of the five role categories of participating classes was calculated in this study. A radar map was applied to illustrate the distribution proportion (Figure 6, Figure 7). In Mag, different classes exhibit varying characteristics of role distribution. First, classes B and A (with ES) had more coordinators and fewer inquirers than Classes C and D (without ES). Second, classes A and C had more integrators and fewer marginals than classes B and D.

Figure 6. The radar map of role distribution of Mag



From Figure 7, the role distribution had some obvious changes after DSR was exchanged. With ES, the role distribution of Class A was mainly similar to before; however, facilitators increased, and inquirers decreased in Class B. Without ES, the inquirers decreased in Class C, while the marginals increased. Additionally, there were more integrators than before in Class D.

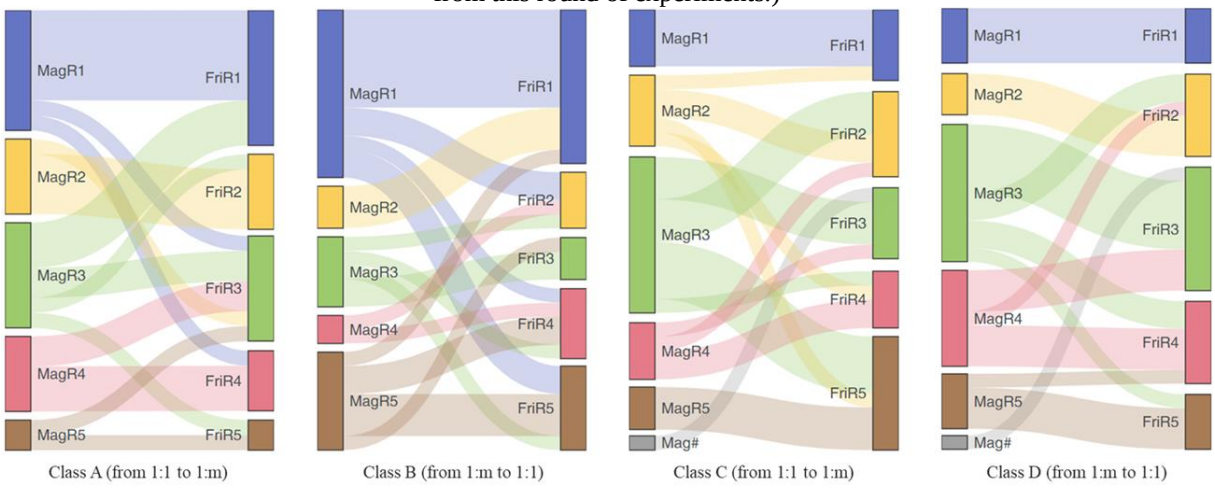
Figure 7. The radar map of role distribution of Fri



4.1.2. Role transition

In inquiry experiments, only the group leader was fixed, and other roles flowed and changed according to the self-regulation of the group. A Sankey diagram was applied to visualize the role transitions of the four participating classes (see Figure 8).

Figure 8. The Sankey diagram of role transition (*Note.* The symbol "#" indicates the students who were absent from this round of experiments.)



As shown in Table 2, when DSR exchanged from 1:1 to 1:m, both Class A and C had a comparatively large number of substitutions of inquirers with other categories. The majority of category transitions in Class A (with ES) were R3 to R1 and R4 to R3, i.e., the work shifted from inquirer to inspector or from facilitator to inquirer. Additionally, most category transitions in Class C (without ES) were R3 to R5 or R2. Most of the inquirers shifted to marginals or integrators, and the role distribution structure changed significantly.

As DSR exchanged from 1:m to 1:1, role transition in class B (with ES) was more frequent, primarily as follows: R1 to R2 or R5; R2 to R1; and R5 to R4. Specifically, some coordinators (group leaders or inspectors) shifted their work focus from coordination to integration, with two inspectors even became spectators. All integrators transformed into coordinators (three recorders were reassigned to inspectors), and two spectators also transformed into facilitators. In Class D (without ES), the transition of roles was relatively sparse compared to

other classes, namely, R4 to R3, and R3 to R2 or R4. Approximately 50% of facilitators shifted to inquirers, while a small portion of inquirers shifted to integrators or facilitators.

In conclusion, the probabilities of role transition were highest for the inquirer (63.64%), and the inquirer may shift to various categories. Second, most transitions were facilitators (50%) and integrators (43.75%). However, the coordinator (the leader is fixed) and marginal categories were relatively stable.

Table 2. Role transition probabilities

	Fri	R1	R2	R3	R4	R5	Total
Mag	R1	**21 (75%)**	2 (7.14%)	1 (3.57%)	2 (7.14%)	2 (7.14%)	28 (100%)
	R2	4 (25%)	**9 (56.25%)**	1 (6.25%)	1 (6.25%)	1 (6.25%)	16 (100%)
	R3	3 (9.09%)	7 (21.21%)	**12 (36.36%)**	4 (12.12%)	7 (21.21%)	33 (100%)
	R4		3 (16.67%)	6 (33.33%)	**9 (50%)**		18 (100%)
	R5	1 (6.25%)		2 (12.5%)	3 (18.75%)	**10 (62.5%)**	16 (100%)
	#			2 (100%)			2 (100%)

Note. The symbol "#" indicates the students who were absent from this round.

4.2. Role interaction

4.2.1. The density and components of group networks

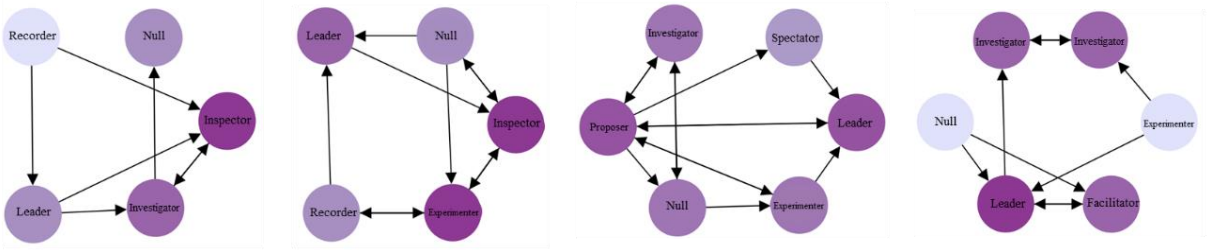
We used Gephi software to map the group's social network (see Figure 9). In the group social network graph, the node contains a role label, and the directed edges represent interactional relations between members. For instance, A→B indicated that A replied in the reflection questionnaire that A and B had more communications. Therefore, the in-degree of a node represented the importance of the corresponding student in the group inquiry. The network density and the action orientation of the participating groups are illustrated in Table 3.

Table 3. The network density and action orientation of groups

Inquiry themes	Mag		Fri	
Class-group	Network density	Action orientation	Network density	Action orientation
A-01	0.35	-1	0.35	-1
A-02	0.40	1	0.42	1
A-03	0.30	1	0.50	1
A-04	0.45	1	0.50	1
A-05	0.45	1	0.50	1
A-06	0.25	-1	0.50	-1
B-01	0.17	0	0.35	1
B-02	0.23	-1	0.33	-1
B-03	0.30	0	0.35	-1
B-04	0.50	1	0.58	1
B-05	0.20	0	0.30	0
B-06	0.60	-1	0.45	0
C-01	0.30	-1	0.40	1
C-02	0.75	1	-	0
C-03	0.43	-1	0.43	0
C-04	0.33	1	0.33	1
C-05	0.30	1	0.50	1
C-06	0.45	-1	0.35	1
D-01	0.60	-1	0.30	1
D-02	0.30	0	0.25	0
D-03	0.60	1	0.58	-1
D-04	0.50	1	0.37	1
D-05	0.45	0	0.45	-1
D-06	0.40	-1	0.25	1

Note. The symbol "-" in Network density represents there are only two students in group C-02 in Fri.

Figure 9. Example graphs representing group social network

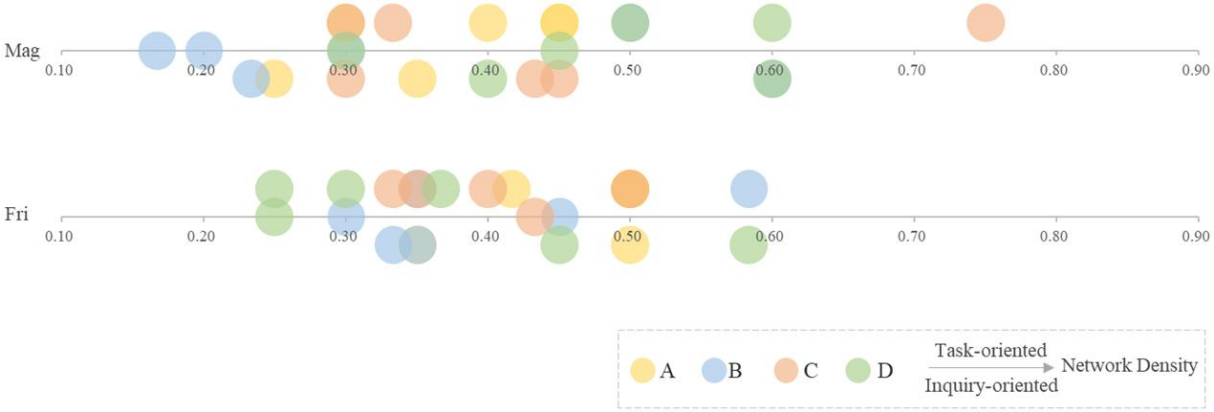


a. Group A-01 in Mag. b. Group B-04 in Mag. c. Group C-03 in Mag. d. Group D-02 in Mag.

4.2.2. The change of role interaction

Figure 10 visualizes the closeness and pattern of role interaction of groups according to Table 3. In Figure 10, dots represent participating groups, and the horizontal axis represents the network density of the group. The dots located above the axis (value was 1) were task-oriented groups; the dots located below the axis (value was -1) were inquiry-oriented groups. In addition, the dots located on the axis (value was 0) were groups without obvious action orientations.

Figure 10. The closeness and pattern of role interaction (*Note.* The transparency of dots is 80%, the color where the dot's color does not match the figure legend indicates that there are more than one dots located in the same position, that is, the network density and the action orientation of the several groups are the same.)



As for network density, in Mag, the ES actually reduced the interaction density of the group, especially in 1:m DSR. The descending order of interaction density was Class D, Class C, Class A, and Class B. When the DSR was exchanged in Fri, the network density of class B was significantly increased. However, the network density of Class D was reduced. In Class A, the network density was also improved, while Class C had no significant changes. This might show that the interaction degree of the group would be improved when the DSR was exchanged under ES. In the matter of action orientation, there was no obvious trend in Mag. However, in Fri, Class C and D had more changes in action orientation than Class A and B. The groups with ES are likely to have a more stable action orientation when the DSR was exchanged.

4.3. Self-efficacy and Collective efficacy

4.3.1. Effects of external scripts

Two separate Jonckheere-Terpstra tests were used to analyze the differences of SE and CE among the four classes (see Table 4). Consequently, there were significant differences of SE and CE in Fri (p = .021, p = .004). Further the Mann-Whitney U test was applied to compare two classes (A and C; B and D; A and B; C and D) in Fri. The results showed that there were significant differences in CE between Class A and C (U = 300.5, z = -2.02, p = .043), and in CE and SE between Class B and D (U = 310, z = -1.991, p = .047; U = 297.5, z = -2.225, p = .026). Additionally, the respective mean values of SE and CE with ES (Class A and B) were higher than those without ES (Class C and D) in the two-round of experiments. This might imply that ES could facilitate students' SE and CE in the long term.

Table 4. Jonckheere-Terpstra test of SE and CE of four classes

		SE		*CE*		*N*
		Mean Rank	*J-T*	Mean Rank	*J-T*	
Mag.	A	69.63	-.806	71.76	-1.380	31
	B	66.71		65.65		34
	C	55.58		59.95		32
	D	66.16		60.68		31
Fri.	A	64.93	-2.301*	69.40	-2.862**	29
	B	67.88		65.12		30
	C	52.95		52.57		29
	D	49.93		48.71		29

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

4.3.2. Changes in SE and CE

To explore the changes in SE and CE before and after the exchange of the DSR, the Wilcoxon signed-rank test was conducted on the two-round of data in the same class. The results showed that there was no significant difference in either SE or CE. However, in Fri, the respective mean values of SE and CE decreased in Class A, C, and D, while the respective mean values of SE and CE increased in Class B.

4.4. Relationships between roles, SE, and CE

The results of the Jonckheere-Terpstra test (see Table 5) revealed a significant difference in SE scores among the five role categories (p = .035). The SE scores of two role categories of all combinations were further compared using the Mann–Whitney U test. There was significant difference in SE between R1 and R5 (U = 876.5, z = -2.276, p = .023). Additionally, the respective mean values of SE and CE of R5 are lower than those of other role categories.

Table 5. Jonckheere-Terpstra test of SE and CE of five role categories

	SE		*CE*		*N*
	Mean rank	*J-T*	Mean rank	*J-T*	
R1	138.31	-2.112*	133.27	-1.335	60
R2	119.79		122.24		43
R3	119.13		114.92		64
R4	129.09		138.79		39
R5	103.26		105.50		39

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

In addition, this study analyzed the scores of group worksheets under four conditions. The Kolmogorov-Smirnov test demonstrated that the data are normally distributed (z = .89, p = .407 > .05). The data are also homoscedastic (F = .536, p = .66 > .05) according to Levene's test. It was found that the DSR had a significant impact on the learning performance (F = 16.596, p < .001) through two-way ANOVA (see Table 6). The learning performance of 1:m groups was significantly better than that of 1:1 groups. However, the ES had no significant impact on the group learning performance.

Table 6. Between-effects of factors on learning performance (N = 48)

	F	Partial η^2
DSR	16.596***	.274
ES	.049	.001
DSR*ES	.049	.001

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

5. Discussion

5.1. Role distribution was affected by DSR and ES

In this study, the concept of role is from the perspective of the task, which is more in line with the collaboration of primary school students (Strijbos & De Laat, 2010). The role reflects the individual's self-awareness about their position in the group. Therefore, role distribution represents how the group works together (Kirschner et al., 2018). By analyzing the difference in role distribution under different conditions, we found that the DSR and the availability of ES affected role distribution interactively. Specifically, when guided by ES, there were more scripted roles, such as leaders and inspectors. According to Strijbos and Weinberger (2010), coordinators play a process-oriented role and facilitate indirect learning by dividing tasks and coordination. Without ES, the collaborative groups were not restricted by external rules but formed internal behavioral patterns by self-regulation (Wang et al., 2017), and more inquirers emerged. However, such differences did not affect the group learning performance.

Under the 1:1 DSR, the operation interface and observation of phenomena were not usually shared in groups, and the knowledge generated during the inquiry process was stored in group members in a distributed manner. Under this constraint, more integrators would be generated. The process of collaboration was described as "inquire by themselves, then discuss" or "everyone did experiment and then drew conclusions" (descriptions in the reflection questionnaire). In contrast, under 1:m DSR, the operation interface was usually shared in real-time. Inquiry activities include collective behavior, as they described it, "we use the tablet together, and discuss the problem together," and "a group with a tablet is easier to discuss." Thus, there were fewer integrators. This can also explain why the learning performance of the group in the 1:m DSR is better than 1:1 DSR. This is consistent with the results of Lin et al. (2012). However, limited by the opportunity to operate VM, it was easy to produce marginal results. According to Simpson et al. (2017), roles with higher status had more control over sharing technological devices, and vice versa. This explained why there were obviously more marginal effects at the 1:m DSR compared with 1:1 DSR.

5.2. Frequent transitions of operational roles in groups emerged when DSR exchanged

How did the DSR exchange impact role distribution? We found that the group carried out internal coordination of roles to adapt to the new resource usage pattern. The division of tasks has changed. The transition of inquirers (R3) was the most obvious, especially from 1:1 to 1:m, which indicated that the ownership of tech tools (Antle, 2014) affected their role perception. For example, a student in Class B wrote that he thought he was "a bystander" because "someone has been operating the tablet" in Mag. However, in Fri, "everyone got a tablet," and he thought of themselves as "explorers." In addition, the role of ES was not clear on role transitions, even though ES may promote the formation of a reasonable role distribution, especially from 1:1 to 1:m (comparison between Class A and Class C). Only in the two-round of experiments could we not see the regular transitions of roles. Perhaps in the longer term, there would be some stable changes according to the development of role knowledge and collaboration skills (Strijbos & Weinberger, 2010).

5.3. ES promoted the stability of role interaction and affected SE and CE

In terms of role interaction, ES played a significant role. On the one hand, it may weaken interaction density in this study even though it showed no effect on group learning performance. This is contrary to Dillenbourg and Jermann (2007); the richness and intensity of interactions between group members determined the effects of collaborative learning. As Vogel and his colleagues (2016) pointed out, the effects of collaboration may be more dependent on the amount of practice of the corresponding activities than on the transitivity of the interaction. On the other hand, when the DSR was exchanged, ES promoted the stability of role interaction, such as more groups having changed action orientation in classes that do not provide ES. When the group has formed a certain interactive mode, the exchange of DSR means a new adaptation. The mean value of students' SE and CE in Class A, C, and D decreased in the second round, indicating that the DSR exchange is a challenge for collaborative learning. In this case, the scaffolding role of the ES could be well-reflected.

At the same time, in Fri, the SE and CE of the scripted classes were significantly higher than that of the unscripted classes. Previous researchers have found that CSCL with ES had a large positive effect on students' collaboration skills and learning outcomes on domain-specific knowledge (Radkowitsch et al., 2020; Vogel et

al., 2016). Furthermore, we found that ES could promote learners' SE and CE in the long term. This may be because ES reduced the difficulties of intragroup coordination, such as division of labor and digression.

5.4. The roles with weaker discourse status had lower SE

Despite the benefits of collaborative learning, uneven participation of roles has been reported in some studies (e.g., Simpson et al., 2017). This study also found that the discourse statuses of the roles were different. The group social network revealed that the interaction of most groups was led by 2-3 roles, generally, the coordinators, integrators, and inquirers, who formed the center of the groups. However, the in-degree of most of the marginal roles ≤ 1. We also found that the SE level of the five role categories were significant different; concretely, the SE of R1 is significantly higher than R5. The status of students in groups is influenced by many factors, such as their interest in exploring the topic and their perceived differences in abilities compared to other group members (Simpson et al., 2017). Additionally, Strijbos and De Laat (2010) pointed out that learners' participation in the collaboration process is influenced by their participation stance, which is related to learning motivation and experience. These differences result in some students doing most of the work and some doing less. In this study, marginals might develop low SE due to a poor collaborative experience. In the reflection questionnaire, we can also feel the loneliness of this type of role: "do nothing," "I can only watch," "soy sauce," "useless guys," (students' descriptions of tasks and roles).

6. Conclusions

This study explored the influence of DSR and ES on CIL. A counterbalanced design of CIL was conducted in four primary school classes. Thematic analysis, social network analysis, and statistical analysis methods were used to analyze the distribution and transition of roles, the interaction between roles, and the SE and CE of the roles. The implication for practice in this study is to help design a collaborative learning environment in science courses. For example, the exchange of DSR can be used to achieve spontaneous role transition, and scripted roles can be designed to promote effective group interaction, such as inspectors. In addition, we need to recognize the marginal roles in collaborative learning and design scaffolds for them. Referring to limitations, 11 students were absent from the second experiment, which impacted group interaction. Moreover, there are influences that are not excluded, such as subject matter. However, this is difficult to avoid when conducting empirical research in authentic classrooms. How to achieve effective and equal collaboration in the role of group members is a problem worthy of further discussion.

Acknowledgement

This project is supported by China Postdoctoral Science Foundation funded project (2021M690433) and the Fundamental Research Funds for the Central Universities (2020NTSS04).

References

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The Ecological dynamics of mathematics education. *The Journal of the Learning Sciences, 25*(2), 203-239.

Antle, A. N. (2014). Scratching the surface: Opportunities and challenges from designing interactive tabletops for learning. In V. R. Lee (Ed.), *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 55-73). Routledge.

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

Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. *International Journal of Science Education, 32*(3), 349-377.

Bouck, E. C., & Flanagan, S. M. (2009). Virtual manipulatives. *Intervention in School and Clinic, 45*(3), 186-191.

Dillenbourg, P., & Evans, M. (2011). Interactive tabletops in education. *International Journal of Computer-Supported Collaborative Learning, 6*(4), 491-514.

Dillenbourg, P., & Hong, F. (2008). The Mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning, 3*(1), 5-23.

Dillenbourg, P., & Jermann, P. (2007). Designing integrative scripts. In F. Fisher, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives* (pp. 275-301). Springer.

Fokides, E., & Mastrokoukou, A. (2018). Results from a study for teaching human body systems to primary school students using tablets. *Contemporary Educational Technology, 9*(2), 154-170.

Gibson, J. J. (1977). The Theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing* (pp. 67-82). Erlbaum.

Girasoli, A. J., & Hannafin, R. D. (2008). Using asynchronous AV communication tools to increase academic self-efficacy. *Computers & Education, 51*(4), 1676-1682.

Goodyear, P., Jones, C., & Thompson, K. (2014). Computer-supported collaborative learning: Instructional approaches, group processes and educational designs. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (4th ed., pp. 439-451). Springer.

Hassler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A Critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning, 32*(2), 139-156.

Hsu, Y. S., & Thomas, R. A. (2002). The Impacts of a web-aided instructional simulation on science learning. *International Journal of Science Education, 24*(9), 955-979.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist, 51*(2), 247-265.

Johnson, D. W., & Johnson, R. T. (1987). *Learning together and alone: Cooperative, competitive, and individualistic learning* (2nd ed). Prentice-Hall.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory Into Practice, 38*(2), 67-73.

Jou, M., Lin, Y.-T., & Tsai, H.-C. (2016). Mobile APP for motivation to learning: An Engineering case. *Interactive Learning Environments, 24*(8), 2048-2057.

King, A. (2007). Scripting collaborative learning processes: A Cognitive perspective. In F. Fisher, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives* (pp. 13-37). Springer.

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning, 13*(2), 213-233.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts – A Conceptual analysis. *Educational Psychology Review, 18*(2), 159-185.

Lin, C. P., Wong, L. H., & Shao, Y. J. (2012). Comparison of 1:1 and 1:m CSCL environment for collaborative concept mapping. *Journal of Computer Assisted Learning, 28*(2), 99-113.

Lloyd, P., & Cohen, E. G. (1999). Peer status in the middle school: A Natural treatment for unequal participation. *Social Psychology of Education: An International Journal, 3*(3), 193-214.

Looi, C.-K., Zhang, B., Chen, W., Seow, P., Chia, G., Norris, C., & Soloway, E. (2011). 1:1 mobile inquiry learning experience for primary science students: A Study of learning effectiveness. *Journal of Computer Assisted Learning, 27*(3), 269-287.

Marin, A., & Wellman, B. (2011). Social Network Analysis: An Introduction. In *The SAGE Handbook of Social Network Analysis* (pp. 11-25). SAGE Publications Ltd.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An Effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education, 96*(1), 21-47.

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

Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A Meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 15*(1), 5-47.

Simpson, A., Bannister, N., & Matthews, G. (2017). Cracking her codes: understanding shared technology resources as positioning artifacts for power and status in CSCL environments. *International Journal of Computer-Supported Collaborative Learning, 12*(3), 221-249.

Strijbos, J.-W., & De Laat, M. F. (2010). Developing the role concept for computer-supported collaborative learning: An Explorative synthesis. *Computers in Human Behavior, 26*(4), 495-505.

Strijbos, J.-W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2007). The Effect of functional roles on perceived group efficiency during computer-supported collaborative learning: A Matter of triangulation. *Computers in Human Behavior, 23*(1), 353-380.

Strijbos, J.-W., & Weinberger, A. (2010). Emerging and scripted roles in computer-supported collaborative learning. *Computers in Human Behavior, 26*(4), 491-494.

Tsai, C. C., Chuang, S. C., Liang, J. C., & Tsai, M. J. (2011). Self-efficacy in internet-based learning environments: A Literature review. *Educational Technology & Society, 14*(4), 222-240.

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2016). Socio-cognitive scaffolding with computer-supported collaboration scripts: A Meta-analysis. *Educational Psychology Review, 29*(3), 477-511.

Vyas, D., Chisalita, C. M., & Van Der Veer, G. C. (2006). Affordance in interaction. In E. Hollnagel (Ed.), *the 13th European Conference on Cognitive Ergonomics: Trust and Control in Complex Socio-technical Systems* (pp. 92-99). ACM.

Wang, C. (2021). Roles interaction during mobile-blended collaborative learning: The Impact of external scripts. In R. Li et al. (Eds.), *Blended Learning: Re-thinking and Re-defining the Learning Process* (pp. 191-213). Springer.

Wang, C., & Le, H. (2021). The More, the merrier? Roles of device-student ratio in collaborative inquiries and its interactions with external scripts and task complexity. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331211010794

Wang, S.-L., & Lin, S. S. J. (2007). The Effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning. *Computers in Human Behavior, 23*(5), 2256-2268.

Wang, S. L., Hsu, H. Y., Lin, S. S. J., & Hwang, G. J. (2014). The Role of group interaction in collective efficacy and CSCL performance. *Educational Technology & Society, 17*(4), 242-254.

Wang, X., Kollar, I., & Stegmann, K. (2017). Adaptable scripting to foster regulation processes and skills in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 12*(2), 153-172.

Wong, L.-H., Chee-Kuen, C., Chee-Lay, T., & Liu, M. (2010). Students' personal and social meaning making in a Chinese idiom mobile learning environment. *Educational Technology & Society, 13*(4), 15-16.

Wong, L.-H., & Looi, C.-K. (2011). What seams do we remove in mobile-assisted seamless learning? A Critical review of the literature. *Computers & Education, 57*(4), 2364-2381.

Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction, 21*(3), 317-331.

Influences of Online Synchronous VR Co-Creation on Behavioral Patterns and Motivation in Knowledge Co-Construction

Hsin-Yun Wang and Jerry Chih-Yuan Sun*
Institute of Education, National Yang Ming Chiao Tung University, Taiwan // wanghsinyun.ie08g@nctu.edu.tw // csun@nctu.edu.tw
*Corresponding author

(Submitted March 24, 2021; Revised August 8, 2021; Accepted August 16, 2021)

ABSTRACT: To explore knowledge co-construction patterns and learning motivation within virtual EFL co-creation environments, this study examined behavioral patterns and motivation in three different co-creation environments (paper-based, 2D digital, and 3D VR co-creation) through sequential behavioral analysis and ANCOVA. The study utilized a quasi-experimental research design with a total of 66 tenth-grade students from two English classes at a public senior high school in northern Taiwan. Based on the visualized behavior transition diagrams, the task-switching behaviors between dissonance identification and knowledge negotiation as well as the isolated behaviors of applying newly-constructed knowledge are the core of knowledge co-construction. Particularly, 3D VR co-creation was characterized by the highest number of higher level isolated acts and lower level circular continuity, both of which reflect VR co-creators' efforts to gain familiarity with advanced technology as well as the intention to exchange information and reach community consensus to overcome task complexity, form community consensus, and lower anxiety. Such behavioral patterns echoed the results of ANCOVA on intrinsic and extrinsic motivation; that is, on either intrinsic or extrinsic motivation, the influence of 3D VR co-creation was the greatest, followed by that of 2D digital co-creation and paper-based co-creation. For future co-creation instruction and research, it is suggested that the instruction of VR co-creation be invested with abundant time to allow mature higher level knowledge co-construction dialogues to occur. Moreover, to gain an even deeper understanding of the social structure embedded in knowledge co-creation, it is suggested that social network analysis (SNA) be employed in future research.

Keywords: Behavioral patterns, Intrinsic motivation, Extrinsic motivation, Co-creation, Virtual Reality

1. Introduction

Social constructivism holds that knowledge is co-created and co-constructed (Mayer, 1996). In this sense, meanings are co-constructed from individuals' jointly-constructed understandings of the world and hence are historically situated in cultural values and practices. With societal contexts and interpersonal relationships constantly shifting, such contextually constructed coordination among people mostly through languages is never fixed and isolated (Camargo-Borges & Rasera, 2013; Gergen & Gergen, 2004); instead, it is fluid, transitory, and dynamic in nature.

Dialogue, creativity and, co-creation may best characterize the post-industrial, social constructivist civilization, where shaping and reshaping meanings and value are not limited to the authorities. Co-creation particularly features today's "network society," where relationship-dominant practice brings co-creativity and conversation for shared leadership and higher equity. Also, it constructs knowledge for common interests or an even higher inter-personal purpose (Camargo-Borges & Rasera, 2013). This generative, participatory fashion later brought about the "learner-as-partner" practice in educational co-creation and knowledge co-construction in the United States, Canada, the United Kingdom, Australia, and Scandinavia (Bovill, 2019).

In the field of education, motivation, defined as the desire to be competent and self-determining in relation to the environment (Deci & Ryan, 1980), is believed to be associated with the four components in learning activities: challenge, curiosity, control, and fantasy (Lepper & Hodell, 1989). Virtual reality (VR), featuring presence, interactivity, and immersion, provides learners with stimuli for make-believe fantasy and curiosity and thus may serve to trigger learners' motivation (Huang et al., 2020; Keller, 2010; Lin et al., 2019; Pintrich, 2003), especially the motivation of foreign language (FL) learners (Lan, 2014).

VR co-creation, the practice of turning the abstract into the concrete through collective consciousness and intelligence in synchronous VR coworking spaces, may further integrate learners' curiosity and a strong sense of community into make-believe activities. Such knowledge co-construction practices may advance to help moderate task difficulty, instead of overwhelming individual students, through learners' improved collaborative

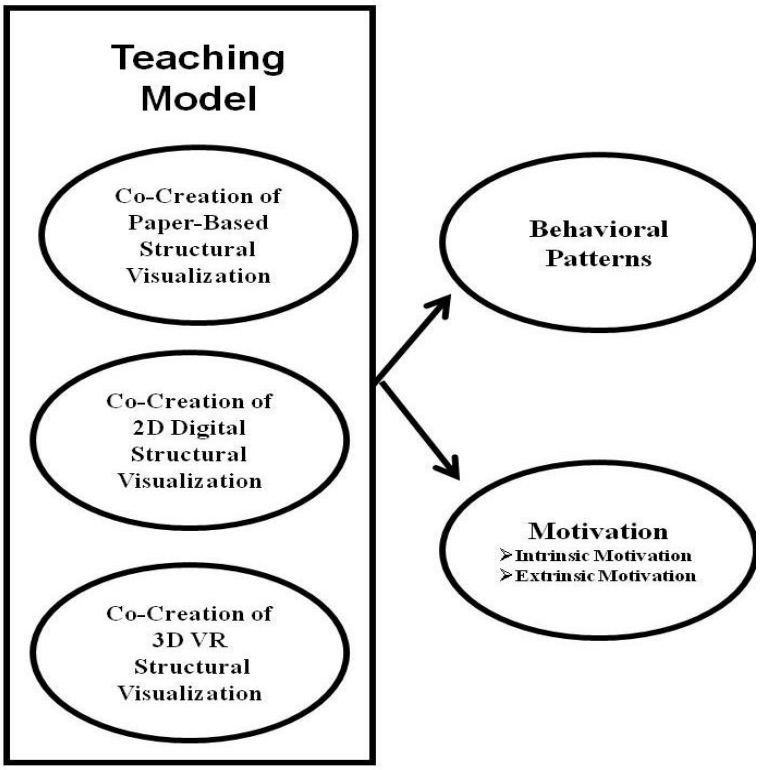
work and learning skills (Blau & Shamir-Inbal, 2017; Bovill et al., 2011; Giner & Peralt Rillo, 2016; Jans et al., 2017) and greater creativity, better communication, and positive group dynamics (Brandt et al., 2008).

Although abundant studies have examined learners' motivation and learning effects in VR, few have visualized the learning paths of language learners in a virtual world (Hsiao et al., 2017). Even fewer have explored the characteristics of students' behavioral patterns and the association between learners' behavioral patterns and their motivation within virtual co-creation environments in an EFL setting. To fill this gap, this study aimed to investigate learners' behavioral patterns and motivation emerging in online synchronous VR co-creation in EFL classrooms in the hope of unveiling and visualizing learners' behavioral sequences and co-creation strategies to better the "learner-as-partner" practice in educational knowledge co-construction. The research model of this study is shown in Figure 1.

To fulfill the abovementioned research objectives, this study addressed the following research questions:

- What are the differences among the sequential behavior patterns in traditional paper-based co-creation, 2D digital co-creation, and 3D VR co-creation for structure visualization?
- What are the differences in motivation among students engaged in traditional paper-based co-creation, 2D digital co-creation, and 3D VR co-creation for structure visualization?

Figure 1. Research model



2. Literature review

2.1. Virtual Reality

VR engages the audience in an immersive, all-inclusive, sensory illusion with presence, interactivity, and immersion (Biocca & Delaney, 1995; Jarvis, 2019). Over the years, VR has been employed in the fields of entertainment (Schlacht et al., 2017), mental health (Maples-Keller et al., 2017), and education (Moro et al., 2017), especially for learning in abstract environments (Thornhill-Miller & Dupont, 2016). For example, Lamb (2014) indicated that VR helps learners not only conceptualize abstract environments that they have never experienced due to physical and technological limitations, but also activate cognitive attributes. Following this thread, Freina and Ott (2015) suggested the effect of the encapsulated stimuli of VR on learners' enhanced retention and efficacy for novel information through interactive life-like experiences. Krokos et al. (2018) also stated that students with VR experiences retain more information and can better apply newly learned knowledge than those without such experiences.

2.2. Learner motivation in VR

Motivation as one of the indispensable components of goal-directed activities shows positive correlations with learners' achievements (Schunk, 2008). According to Lepper et al. (2005), motivation is categorized into intrinsic and extrinsic motivation; the former refers to the motivation of engaging in one activity for its own sake due to its inherently interesting and enjoyable nature, whereas the latter impels individuals to become involved for explicit outcomes or avoidance of failure and punishment. To assess learners' motivation, numerous studies have employed various educational technologies, such as clicker technology and mobile polling (Sun, 2014), integrated concept maps and classroom polling systems (Sun, et al., 2018a), challenging games (Sun, 2018) and gamified interactive response system (IRS) (Sun & Hsieh, 2018) in their curriculum designs. Moreover, numerous researchers have targeted the association of motivation and learners' behavioral patterns via lag sequential analysis. For example, Sun et al. (2017) found that game-based learning triggered learners' persistent "learning with gaming" behavioral pattern for better knowledge retention. Sun, et al. (2018a) examined the behavioral sequences in a votable concept mapping approach. Also, Sun, et al. (2018b) explored highly motivated students' serious reading patterns and suggested that online reading duration is a significant indicator of online reading motivation.

Further, many studies have claimed the positive relationship between intrinsic motivation and VR (Keller, 2010; Lin et al., 2019). For example, Dalgarno and Lee (2010) observed that spatial knowledge was enhanced and intrinsic motivation increased due to the contextualization of VR. Lin et al. (2019) examined students' situational interest in VR-guide and map-guide groups, indicating the impact of VR guidance on yielding situational interest and motivation. Similarly, Huang et al. (2020) and Linnenbrink and Pintrich (2019) stated that VR presents strong situativity to stimulate learners' intrinsic motivation as well as attention to learning.

2.3. Digital co-creation of VR

Co-creation, different from collaboration, refers to collective wisdom and efforts invested by professionals and nonprofessionals coexisting in communities without clearly specified accountability or leadership duties. Recent studies have shown positive relationships between co-creation and learners' self-awareness of their learning progress and subject knowledge (Elsharnouby, 2015; Lubicz-Nawrocka, 2018), improved collaborative work and learning skills (Blau & Shamir-Inbal, 2017; Jans et al., 2017), increased satisfaction, trust, and loyalty to groups (Blau & Shamir-Inbal, 2017; Giner & Peralt Rillo, 2016), and greater creativity, better communication and positive group dynamics (Brandt et al., 2008). Many studies have particularly indicated the influential impacts of equity commonly shared in digital co-creation on students' learning enthusiasm and motivation (Bovill, 2019; Yilmaz et al., 2020), yet potentially accompanied by inefficiency owing to the higher need for negotiation and compromise (Wang & Sun, 2021; Yilmaz et al., 2020).

VR creation, from the perspective of social constructivism, stimulates collaboration and situativity in knowledge development through diversified and authentic contexts. For example, Grover et al. (2015) indicated the benefits of VR creation for facilitating critical thinking and problem-solving during VR collaborative creation. Castaneda et al. (2017) stressed the contribution of VR creation to meeting learners' diversified learning needs. Yeh et al. (2018) further determined elementary school students' enhanced autonomy and gender differences in collaborative VR creation of authentic contexts. Particularly, Lan (2014) further indicated how VR facilitated effective foreign language (FL) learning through immersion, interaction, and authenticity. Broadbent and Poon (2015) conducted a content analysis to investigate the state of VR and the learning gains in the VR-supported language learning process. Hsiao et al. (2017) explored the learning strategies and approaches used in a virtual world via visualization analytics. Unveiling the learning benefits VR might bring, Koch et al. (2018) yet further indicated the necessity of extending course duration for co-creators to overcome the novelty effect for improved performances and opportunities to challenge higher levels of knowledge co-construction and co-production.

Given an even higher degree of equity and flexibility in task accountability, VR co-creation has further encouraged collective consciousness among learners in web-based real-time co-working platforms. For example, Google CoSpaces Edu is equipped with a user-friendly interface and coauthoring tools for VR co-creators to simultaneously co-edit texts and co-create VR scenes within a virtual environment. Moreover, its coding feature equips learners with little or no programming skills to visualize the abstract by programming self-created objects to follow instructions. GmbH (2017) considered CoSpaces a complement to traditional teaching methods by immersing students in a VR world where coding can be connected with the curriculum on a completely new level. Recent studies (Bertolini et al., 2018; Krause, 2017) have also shown positive learning outcomes of Google CoSpaces Edu in primary and secondary education; students using CoSpaces for digital storytelling to

co-create virtual art exhibitions (Bertolini et al., 2018) and historical scenes (Krause, 2017) had better comprehension and retention of subject-matter knowledge.

Ideally expected to stimulate autonomy, collaboration, and cognitive learning, the question of whether 3D VR co-creation consists of distinct knowledge construction behavior remains unexplored. To be more specific, little research has examined real-time VR co-creation in terms of sequential behavioral patterns and motivation in an EFL setting. To fill the gap, this study explored learners' collective intelligence and consciousness from the perspective of behavioral patterns and motivation in a VR co-creation project in EFL classrooms.

3. Methods

The study explored learners' behavioral patterns and assessed their motivation in a VR co-creation project in EFL classrooms. To achieve these objectives, information about the participants, methods, instructional design, and study instruments, an overview of the co-creation platforms, and the data collection and analysis procedures are provided and illustrated in the following sections.

3.1. Participants

This quasi-experimental research was conducted in 2021 and involved one teacher and a total of 66 10th-grade students, including 39 males (59%) and 27 females (41%) from two English classes at a public senior high school in northern Taiwan. The average age of the students was 16.47 years (*SD* = .56). To assess the effects of the paper-based, 2D Jamboard, and VR CoSpaces environments on co-creation, the participants were randomly assigned to a control group, experimental group A, and experimental group B, with a valid sample of $n = 22$ in each group. It is noteworthy that the subgroup gender was not as balanced as the overall student body, as shown in Table 1, due to the random sampling policy, which was used to help create an authentic context of co-creation in the classrooms.

Table 1. The gender ratio of the three subgroups

	Male (34 students)	Female (32 students)
Paper-based Co-creation (22 students)	7 (32%)	15 (68%)
2D Jamboard Co-creation (22 students)	11 (50%)	11 (50%)
3D VR CoSpaces Co-creation (22 students)	16 (73%)	6 (27%)

3.2. Methods and instructional design

The experimental process of this study is shown in Figure 2. Session 1 (100 minutes) involved a pre-test on intrinsic and extrinsic motivation, along with the training on tool use and the reading strategies for the two genres, namely descriptive narratives and expository announcement, through which the students were introduced to Kinkaku-ji in Kyoto during the COVID-19 pandemic. To be more specific, the descriptive narrative, entitled "Kyoto: The Heart of Japan," was required to be rewritten with reference to the expository announcement, entitled "Sightseeing Facilities in Japan Reopening with Preventative Measures for COVID-19." To meet the goal, several comprehension strategies for long passage reading such as scanning, skimming, and graphic organizers were taught to organize and visualize ideas, which, based on Lan (2013), may facilitate vocabulary learning while reading.

Session 2 (200 minutes) was the phase of co-creation, where the student co-creators were engaged in the idea visualization tasks for the abovementioned genre reading and writing mission; the control group was engaged in paper-based co-creation, while experimental groups A and B used Google Jamboard and CoSpaces for digital 2D and 3D VR co-creation, respectively (see Figure 3). The process of co-creation was video- and screen-recorded for the sequential behavioral analysis. The experiment ended in Session 3 with post-tests on intrinsic and extrinsic motivation.

Figure 2. The experimental flow

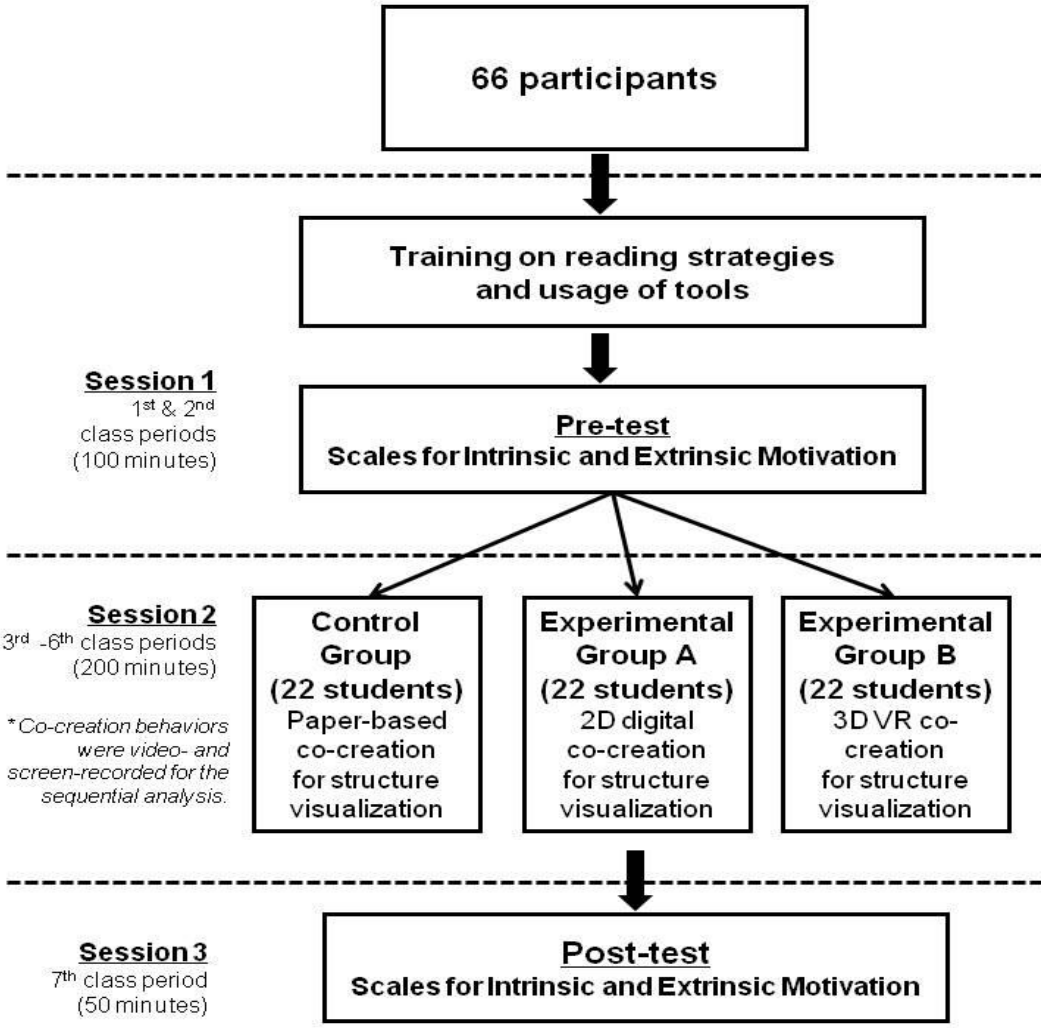
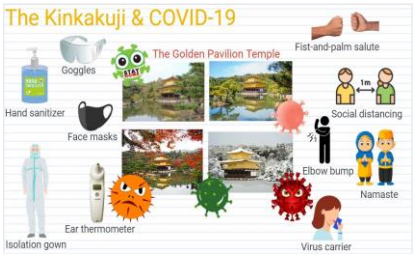


Figure 3. Co-creation interfaces of the various groups

Control Group



Experimental Group A

Experimental Group B

3.3. Instruments

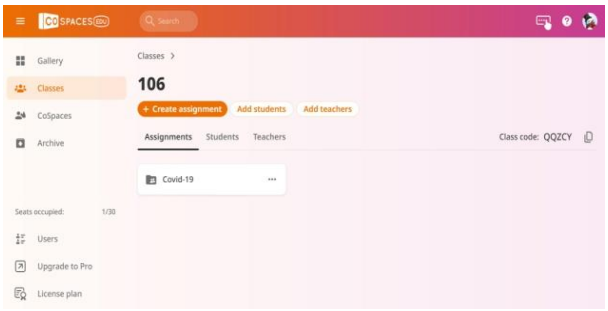
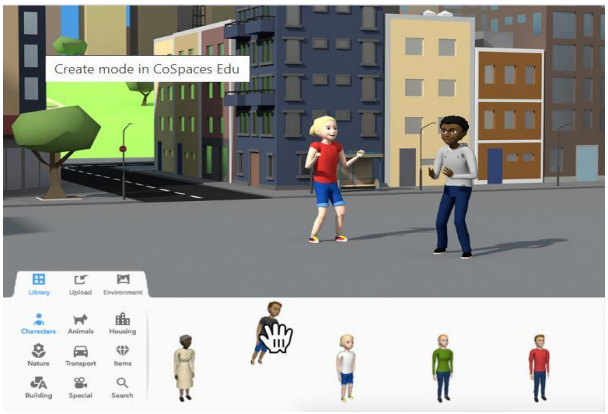
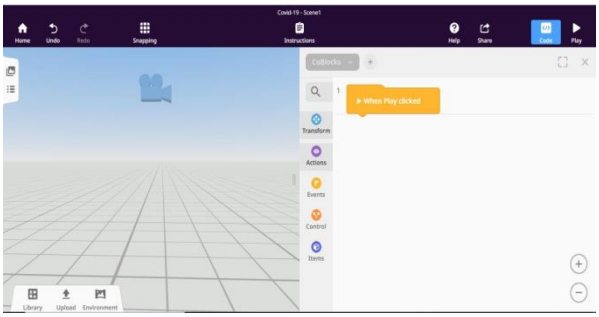
The scale assessing motivation, which was based on the scale of Pintrich et al. (1991), with reference to that of Sun and Hsieh (2018), was a 6-point Likert scale with four questions on intrinsic motivation and four questions on extrinsic motivation. An example question for intrinsic motivation is, “In a class like this, I prefer course material that challenges me so I can learn new things.” An example question for extrinsic motivation is, “Getting a good grade in this class is the most satisfying thing for me right now.” In terms of the reliability of the post-test, the Cronbach’s alpha value for the scale was .94, while the reliability values of the constructs ranged from .89–.94, indicating excellent overall internal consistency (George & Mallery, 2003).

3.4. Digital co-creation platforms: Google Jamboard and CoSpaces

The digital co-creation platforms were Google Jamboard and CoSpaces. The former was used for 2D structure visualization co-creation, whereas the latter functioned as a co-working space for learners to create context-based simulated scenarios that could be explored virtually using cardboard headsets. Both Google Jamboard and CoSpaces can be co-authored, allowing users to simultaneously co-edit texts and co-create objects and scenes within a virtual environment. Particularly, CoSpaces enables users to create 3D VR scenarios for use as backgrounds. Learners may further make objects interactive through Scratch-like coding within CoSpaces. The features of Google CoSpaces Edu are listed and illustrated in Figure 4.

After co-designing, co-constructing, and co-coding their virtual worlds, students can explore them using a Google Cardboard headset or similar VR viewers. Finally, users may share their fully immersive worlds with others via a link or QR code.

Figure 4. Features of Google CoSpaces

Feature	Description	Illustration
Dashboard	On the dashboard are the gallery and the control panel for assigning students to groups.	
Creation Toolbox	The creation toolbox includes assets and buildings for creative co-creation.	
Coding Features	Brick-like coding for the programming of characters/objects to follow instructions.	

3.5. Data collection and analysis

A large amount of video-based data was used for the sequential analyses. Firstly, the entire process, involving a total of 66 students and more than 150 hours of video data, was screen- and video-recorded. Secondly, the

recorded co-creation behaviors were chronologically coded in 5-second slots based on the coding scheme of Social Construction of Knowledge (see Table 2), based on the Interaction Analysis Model (IAM) proposed by Gunawardena et al. (1997) to qualitatively examine types of cognitive activities performed by co-creators through their behaviors of questioning, clarifying, negotiating, and synthesizing during co-creation.

Two skilled coders with educational backgrounds were recruited to ensure the consistency, reliability, and validity of the data. To examine the reliability of the data, the complete video data of one sampled student was firstly encoded in the pilot study by both coders. The analysis revealed the reliability of the Multi-rater Fleiss Kappa coefficient k to be .63, which achieved good consistency (Landis & Koch, 1977). Finally, a total of 94,776 behavioral codes were yielded.

Table 2. Interaction analysis model for examining knowledge construction in co-creation activities

Coding	Behavior	Definition	Examples
C1	Comparing of Information	A statement of observation or opinion	Co-creators sharing observations and opinions.
		A statement of agreement from one or more other participants	Co-creators agreeing with each other.
		Corroborating examples provided by one or more participants	Co-creators supporting each other by strengthening provided examples.
		Asking and answering questions to clarify details of statements	Co-creators asking and answering each other to gain a deeper understanding.
		Definition, description, or identification of a problem	Co-creators providing definitions, descriptions, or identification of a problem.
C2	The Discovery and Exploration of Dissonance or Inconsistency Among Ideas, Concepts or Statements	Identifying and stating areas of disagreement	Co-creators indicating cognitive incongruity.
		Asking and answering questions to clarify the source and extent of disagreement	Students asking and answering each other to identify cognitive dissonance.
		Restating the participant's position, and possibly advancing arguments or considerations in its support by references to the participant's experiences, literature, formal data collected, or proposal of relevant metaphors or analogies to illustrate points of view	Co-creators strengthening statements with further information or data, whether formal or not.
C3	Negotiation of Meaning/Co-Construction of Knowledge	Negotiation or clarification of the meaning of terms	Co-creators clarifying or negotiating over confusing expressions.
		Identification of areas of argument or overlap among conflicting concepts	Co-creators comparing and contrasting concepts.
		Proposal and negotiation of new statements embodying compromise and co-construction.	Co-creators proposing and negotiating over new ideas.
		Proposal of integrating or accommodating metaphors or analogies	Co-creators proposing to integrate and accommodate metaphors or analogies.
		Testing the proposed synthesis against "received fact" as shared by the participants and/or their culture	Co-creators challenging the "received fact."
C4	Testing and Modification of Proposed Synthesis or Co-Construction	Testing against an existing cognitive schema	Co-creators testing a new method against an existing cognitive schema.
		Testing against personal experience	Co-creators testing a new method against personal experience.
		Testing against formal collected data.	Co-creators testing a new

			method against formal collected data.
		Testing against contradictory testimony in the literature	Co-creators testing a new method against existing literature.
		Summarization of agreement(s)	Co-creators summarizing consensus.
C5	Agreement Statement(s)/Applications of Newly-Constructed Meaning	Application of new knowledge	Co-creators applying their co-constructed knowledge.
		Meta-cognitive statement by the participants illustrating their understanding that their knowledge or ways of thinking (cognitive scheme) have changed as a result of the conference interaction	Co-creators expressing their changed understanding after knowledge co-construction.

4. Results and discussion

To answer the first research question on the sequential behavioral patterns in the various co-creation environments, the data coded based on the Interaction Analysis Model (IAM) proposed by Gunawardena et al. (1997) underwent sequential analysis. Firstly, the adjusted residuals tables of the three groups as shown in Tables 3, 4, and 5 were generated and are elucidated in the following sections. The z-score values were calculated to determine the continuity of each sequence; as Bakeman and Gottman (1997) indicated, a z-score greater than +1.96 implies the significance ($p < 0.05$) of a sequence. Based on the z-score values, the diagrams of the behavioral transition patterns were visualized, as shown in Figures 5, 6, and 7.

To answer the second research question on motivation in the three co-creation environments, the analysis of covariance (ANCOVA) in SPSS 20 was performed to identify between-group differences, with the pre-test as the covariant, the post-test as the dependent variable, and the co-creation mode as the fixed factor.

4.1. The sequential patterns of the paper-based co-creation

The adjusted residuals of the behavioral transition patterns in paper-based co-creation are presented in Table 3. The z-score values were visualized in the transition diagram, as shown in Figure 5, where the arrows disclosed the directions of the seven sets of significant sequences, including isolated continuity (C1→C1 and C5→C5), unidirectional movements (C4→C2) and bidirectional movements (C2←→C3 and C3←→C4).

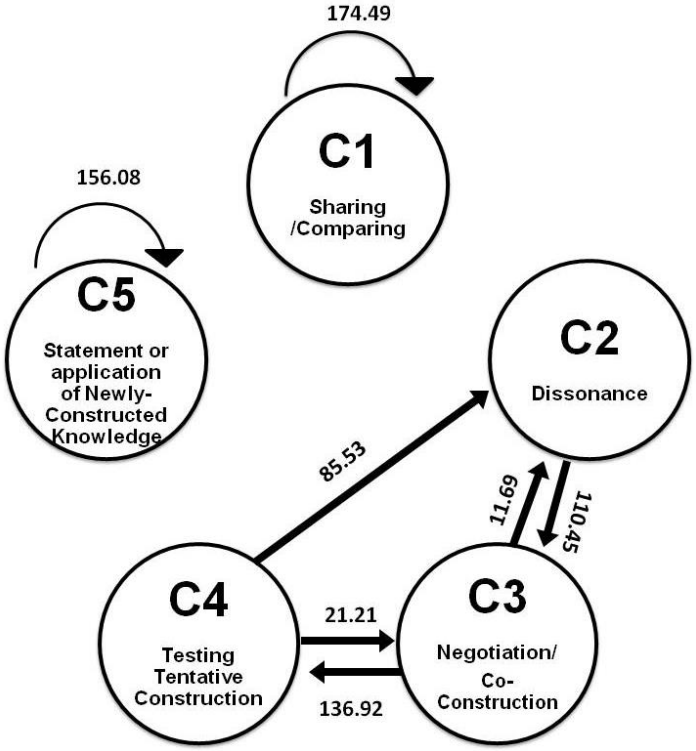
It is noteworthy that both the isolated continuity of C1→C1 and C5→C5 reached a high level of significance; the former reached the z-score value of 174.49 ($p < .05$) while that of the latter was 156.08 ($p < .05$). As for the bidirectional continuity, the z-score values of the forward movements from C2 to C3 ($z = 110.45$, $p < .05$) and from C3 to C4 ($z = 136.92$, $p < .05$) showed greater significance than those of the backward movements from C3 to C2 ($z = 11.69$, $p < .05$) and from C4 to C3 ($z = 21.21$, $p < .05$). Furthermore, only one unidirectional sequence of C4→C2 was discovered ($z = 85.53$, $p < .05$), with which learners in paper-based co-creation made it possible to form a recurring cycle across C2, C3, and C4. Throughout such a circular behavioral cycle, paper-based co-creators might frequently and repeatedly identify disagreement, co-construct knowledge, and test tentative knowledge structures.

Table 3. The results of sequential analysis of behaviors in the paper-based co-creation

	C1	C2	C3	C4	C5
C1	174.49*	-13.32	-30.58	-23.56	-43.93
C2	-18.66	-22.29	110.45*	-27.63	-51.51
C3	-30.03	11.69*	-12.23	136.92*	-82.89
C4	-23.13	83.53*	21.21*	-34.25	-33.66
C5	-42.87	-51.21	-58.90	-63.48	-42.87

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

Figure 5. Sequential patterns of the paper-based co-creation



4.2. The sequential patterns of the 2D Jamboard Co-Creation

The adjusted residuals of the 2D Jamboard Co-Creation are presented in Table 4, where the z-score values were visualized in the diagram of the behavioral transition patterns, as shown in Figure 6. In this diagram, it is noticed that the arrows disclose the directions of the eight sets of significant sequences, including isolated continuity (C3→C3, C4→C4, and C5→C5), unidirectional movements (C3→C4) and bidirectional movements (C1←→C2 and C2←→C3).

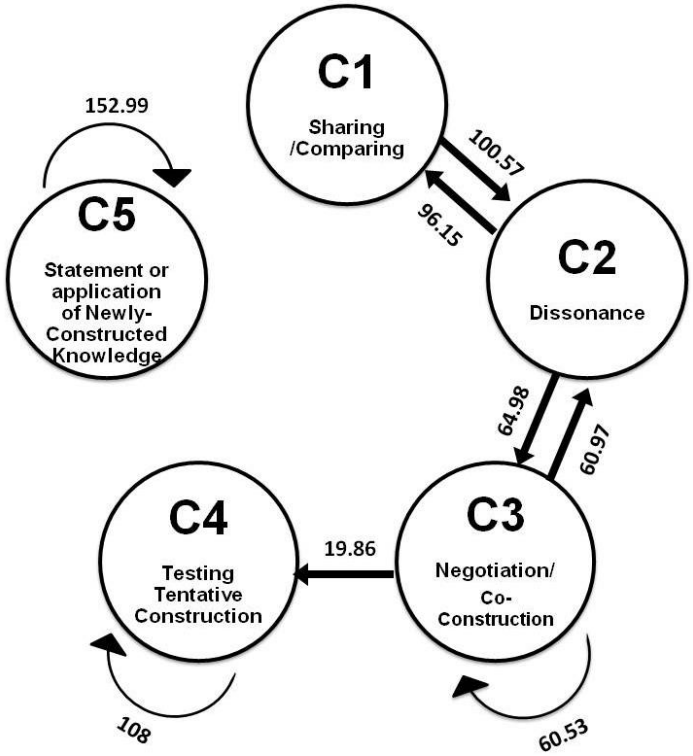
It is noteworthy that the isolated continuity of C5→C5 (z = 152.99, p < .05) and C4→C4 (z = 108, p < .05) showed much higher significance than that of C3→C3 (z = 60.53, p < .05). Moreover, in the bidirectional continuity, the z-score values of the forward movements from C1 to C2 (z = 100.57.45, p < .05) and from C2 to C3 (z = 64.98, p < .05) were rather close to those of the backward movements from C2 to C1 (z = 96.15, p < .05) and from C3 to C2 (z = 60.97, p < .05), which displayed a contrast to the bidirectional movements in paper-based co-creation. Such contrast might result from the digital co-creators' access to abundant online resources and their greater needs for examination and re-examination of the information sent and received if compared with paper-based co-creators working face-to-face on hands-on materials. Furthermore, only one unidirectional sequence of C3→C4 was discovered with a considerably low z-value of 19.86 (p < .05). In short, unlike paper-based co-creation, there existed no frequent task-switching behaviors or circular behavioral cycles beyond C3 in the 2D digital co-creation. Instead, 2D co-creators depended more on specific isolated behaviors and encountered greater challenges of advancing to higher levels of knowledge co-construction, which might potentially cause insufficiency, as indicated by Wang and Sun (2021) and Yilmaz et al. (2020).

Table 4. The results of sequential analysis of behaviors in the 2D Jamboard co-creation

	C1	C2	C3	C4	C5
C1	-5.88	100.57*	-17.70	-13.91	-32.96
C2	96.15*	-19.20	64.98*	-24.58	-58.26
C3	-16.92	60.97*	60.53*	19.86*	-94.91
C4	-13.30	-24.58	-23.41	108.00*	-38.57
C5	-31.34	-57.93	-65.83	-74.17	152.99*

Note. *p < .05.

Figure 6. Sequential patterns of the 2D Jamboard co-creation



4.3. The sequential patterns of the 3D VR Co-Creation

The adjusted residuals of the sequential patterns of the 3D VR co-creation are presented in Table 5. The z-score values were visualized in the diagram of the 3D VR behavioral transition patterns, as shown in Figure 7. In this diagram, it is noticed that the arrows disclose the directions of the eight sets of significant sequences, including isolated continuity (C2→C2, C3→C3, C4→C4, and C5→C5), unidirectional movements (C2→C3 and C3→C1), and bidirectional movements (C1←→C2).

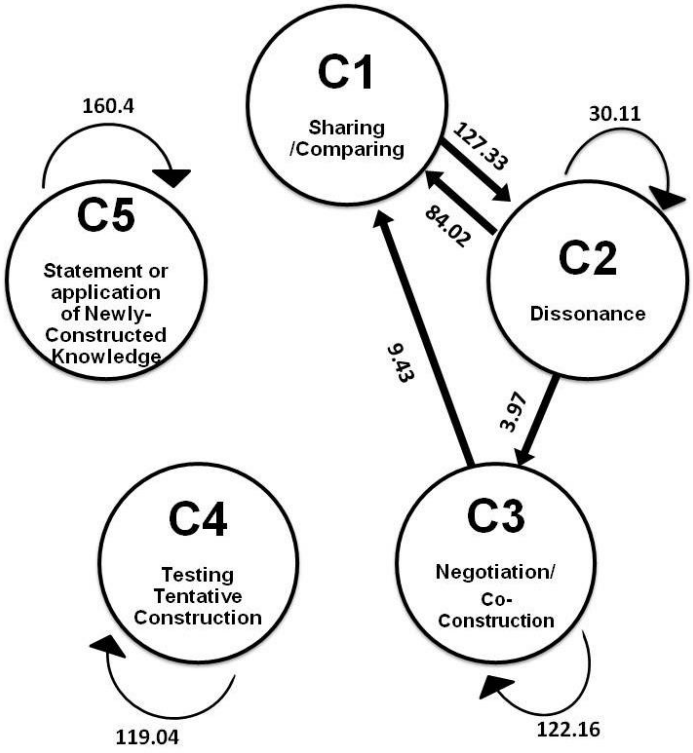
Compared with paper-based and 2D digital co-creation, in 3D VR co-creation there existed the most isolated continuity and the least number of bidirectional movements. Specifically, isolated continuity included C5→C5 (z = 160.4, $p < .05$), C4→C4 (z = 119.04, $p < .05$), C3→C3 (z = 122.16, $p < .05$), and C2→C2 (z = 30.11, $p < .05$), whereas bidirectional continuity fell on C1←→C2 with z-values of 127.33 and 84.02 ($p < .05$). Furthermore, two comparatively weak unidirectional sequences of C2→C3 (z = 3.97, $p < .05$) and C3→C1 (z = 9.43, $p < .05$) were discovered to form a circular behavioral pattern across C1, C2, and C3. These findings in VR co-creation support the novelty effect whereby the more advanced technology is, the more efforts co-creators have to exert on specific isolated behaviors, by which they attempt to improve their performance and to challenge higher levels of knowledge co-construction and co-production (Koch et al., 2018).

Table 5. The results of sequential analysis of behaviors in the 3D VR co-creation

	C1	C2	C3	C4	C5
C1	-27.10	127.33*	-34.97	-26.43	-47.02
C2	84.02*	30.11*	3.97*	-36.90	-65.63
C3	9.43*	-48.82	122.16*	-11.30	-59.95
C4	-26.12	-19.17	-33.70	119.04*	-19.49
C5	-46.15	-65.21	-59.56	-22.37	160.40*

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

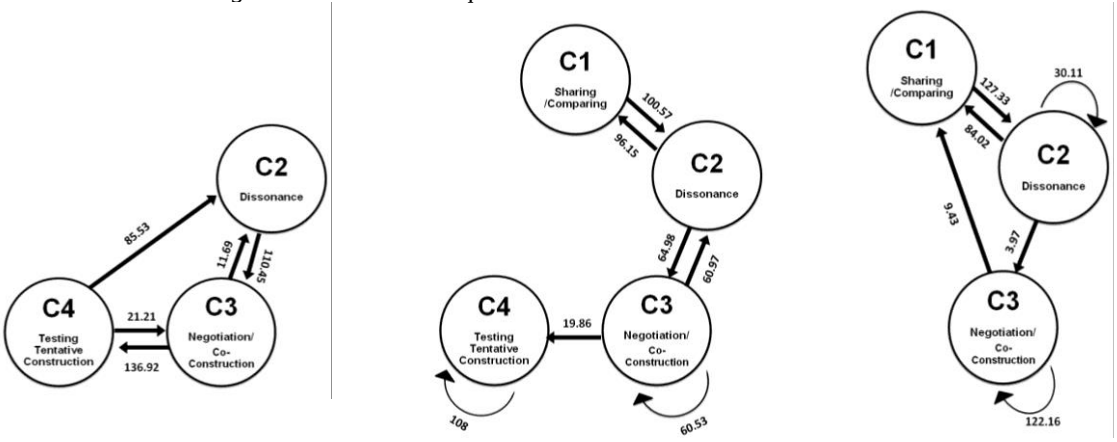
Figure 7. Sequential patterns of the 3D VR co-creation



4.4. Associations between co-creation environments and knowledge construction behaviors

Figures 8 and 9 highlight the behavioral sequences and isolated behavior contiguity in the three various co-creation groups.

Figure 8. Behavioral sequences in the three co-creation environments



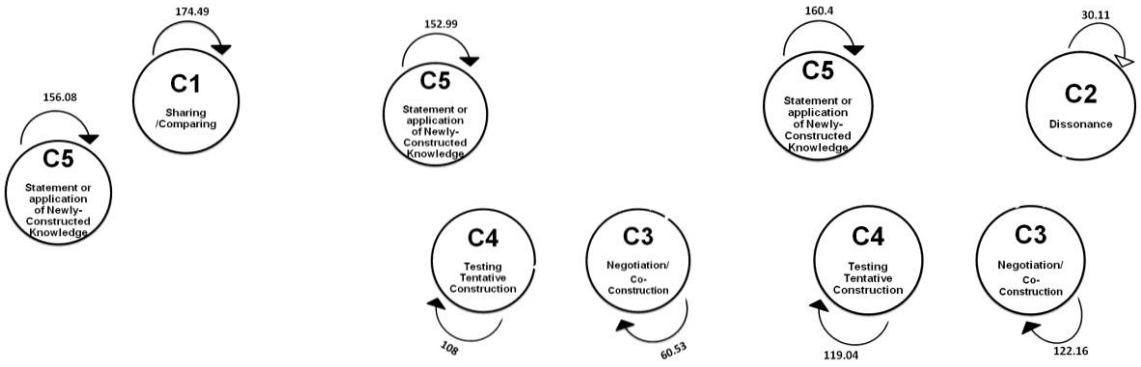
Control Group (Paper-based) | Experimental Group A (2D Jamboard) | Experimental Group B (3D VR CoSpaces)

Figure 8 is the visualized diagram highlighting the chained behaviors in the paper-based, 2D digital, and 3D VR groups. The commonality among the three is the sequence of C2→C3, suggesting that the behavioral movement from dissonance identification to meaning negotiation is the core of knowledge co-construction. On the other hand, it is noteworthy that the pattern between C2 and C3 distinguished the 2D and 3D VR co-creation behaviors. To be more specific, in 2D digital co-creation, C2→C3 was paired with C3→C2 to form bidirectional movements, whereas in VR co-creation there existed the unidirectional sequence of C2→C3 with a weak p-value; yet it was accompanied by strong but isolated continuity of C2 and C3. In other words, VR co-creators were stuck in repeated isolated behaviors (e.g., dissonance identification, knowledge negotiation) in the hope of reaching consensus with peers and gaining familiarity with the new technology, which, as Koch et al. (2018)

indicated, does indeed reflect the novelty effect of new technology and may, in turn, bring about the learners' persistence in specific isolated learning behaviors.

Additionally, the circular chained sequences of behavioral movements were observed in VR co-creation, but not in 2D digital co-creation. This indicates that the learners in the VR co-creation mode tended to return to the lower knowledge co-construction levels from the higher-level ones. Moreover, VR co-creators' stronger dependence on information sharing and comparing (C1) and greater difficulty in advancing to testing tentative knowledge structures (C4) due to the higher complexity of VR co-creation further distinguished their circular sequences from those of 2D digital co-creators.

Figure 9. Isolated continuity in the three co-creation environments



Control Group (Paper-and-pen) | Experimental Group A (2D Jamboard) | Experimental Group B (3D VR CoSpaces)

Figure 9 highlights the isolated contiguity in co-creation. As shown, the co-creators in the three different environments all showed fervent effort in C5, indicating learners' great motivation to apply newly constructed knowledge for task completion no matter which co-creation group they belonged to.

Moreover, it should be noted that in VR co-creation there existed the highest number of instances of isolated continuity, manifesting a great contrast to the least number in paper-based co-creation. On the other hand, the isolated C2 and C3 continuity featured not only VR co-creators' persistent needs for information sharing but stronger eagerness for cognitive dissonance resolution, if compared with those of the 2D digital co-creators. These support the previously mentioned claim that VR co-creators tended to repeat specific isolated behaviors to gain familiarity with new technology. Also, owing to the complexity of digital tool use, VR co-creators depended more on basic knowledge co-construction acts and encountered greater obstacles in advancing to higher hierarchical behaviors. Both the findings echoed the claim of Koch et al. (2018) that the novelty effect leads to higher motivation to perform particular isolated behaviors to get better control over new technology and to improve performance.

4.5. Motivation in different co-creation environments

ANCOVA in SPSS 20 was performed for student motivation to identify between-group differences, with the pre-test as the covariant, the post-test as the dependent variable, and the co-creation mode as the fixed factor. With the insignificant result of the homogeneity test, the ANCOVA results showed that the effect of interaction between the covariates and variables was significant ($F = 1823.06$, $p < .00$), as was the effect of the co-creation environments on student motivation ($F = 2409.31$, $p < .00$).

Table 6. Summary of covariance analysis for motivation

Source of variance	*SS*	*df*	*MS*	*F*	*p*	Partial η^2
Covariates (pre-test on motivation)	30.86	1	30.86	1823.06	.00	96.7%
Inter-group (co-creation platforms)	81.58	2	40.79	2409.31	.00	98.7%
Intra-group (Error)	1.06	62	.02			
Overall	1019.82	66				

As shown in Table 6, the effects of the pre-test and the co-creation platforms were significant on the learners' overall motivation. Specifically, the covariate significantly predicted the dependent variable ($F = 1823.06$, p

< .00), suggesting that the post-test on motivation was influenced by the pre-test. Moreover, with the pre-test effect removed, the effect of the co-creation environments was still significant (F = 2409.31, p < .00). In other words, the post-test motivation score was influenced by the co-creation environments.

Further ANCOVA results on the two constructs of motivation (intrinsic and extrinsic motivation) are reported in Table 7 below. In terms of intrinsic motivation, the effect of the co-creation environment was significant (F = 1282.06, p < .00). The effect of the co-creation environment was also observed to be significant for extrinsic motivation (F = 1024.57, p < .00). In other words, both the post-test intrinsic and extrinsic scores were influenced by the co-creation environments. Particularly, on either intrinsic or extrinsic motivation, the influence of 3D VR co-creation was greater than that of 2D digital co-creation, and both were greater than that of paper-based co-creation as a means of structure visualization.

Table 7. Summary of the covariance analysis of the constructs of motivation

Constructs	*F*	*p*	Post hoc
Intrinsic Motivation	1282.06	.00	(3) > (2) > (1)
Extrinsic Motivation	1024.57	.00	(3) > (2) > (1)

Note. (1) = Control Group; (2) = Experimental Group A; (3) = Experimental Group B.

4.6. Associations between behavioral patterns and motivation in co-creation environments

Based on the aforementioned findings of the sequential behavioral analysis, the high frequency of C2→C3 and the isolated congruity of C5 among all the groups were noticed, connoting not only the student co-creators' persistent efforts on idea negotiation and structure construction, but also their motivating enthusiasm of completing co-creation tasks whichever co-creation group they belonged to. Such behavioral patterns could be further reflected by the ANCOVA results that the overall motivation was significantly influenced by the various co-creation platforms, which is consistent with the findings of Brandt et al. (2008), Hwang and Chen (2017), Bovill (2019), and Yilmaz et al. (2020) that co-creation with various levels of equity may influence learners' group dynamics and learning motivation.

On the other hand, compared with 2D and paper-based co-creators, the preference of 3D VR co-creators for lower-level task-switching behaviors and isolated higher-level behaviors not only echoed the post-hoc results of ANCOVA that the influence of 3D VR co-creation on either intrinsic or extrinsic motivation was greater than that of 2D digital and of paper-based co-creation, but also explained the two sides of one coin: the avoidance of project failure and uncertainty over advanced technology on one side, and the motivating enthusiasm for visualizing such make-believe fantasy on the other side. While the former manifested 3D VR co-creators' extrinsic motivation as a means of reward or avoidance of failure, the latter highlighted their intrinsic motivation as the driving force to accomplish achievements due to inherent interest or ambition (Schunk, 2008). Such intrinsic motivation, based on Dalgarno and Lee (2010), GmbH (2017), Linnenbrink and Pintrich (2019), and Huang et al. (2020) may particularly result from the highly immersive virtual environment in VR co-creation, and may function as a complement to traditional teaching methods for students to visualize the abstract via co-creativity and knowledge co-construction (Lan, 2020; Lepper et al., 2005; Lepper & Hodell, 1989). Yet, such extrinsic motivation also served as a reminder for instructors to stress the importance of preparatory training on language learning strategies in virtual worlds if learning motivation and outcomes are expected (Lan, 2020).

5. Conclusion and implications

The purpose of this study was to investigate learners' behavioral patterns and motivation of online synchronous VR co-creation in EFL classrooms through sequential behavioral analysis and ANCOVA. Based on the z-score values generated in the sequential analysis, the visualized diagrams of the behavioral transition patterns revealed the fact that the behaviors of dissonance identification and knowledge negotiation were the core of the knowledge co-construction process. Moreover, the persistent isolated behavior of applying newly-constructed knowledge, as shown in every co-creation group, highlighted the learners' enthusiasm and motivation for co-creation, as indicated by Brandt et al. (2008).

The visualized transition diagrams also showed that VR co-creation involved the most isolated behavioral continuity, whereas paper-based co-creation showed the least. This reflected VR co-creators' persistence in repeating isolated learning behaviors to gain higher levels of familiarity with new technology, as Koch et al. (2018) indicated that the novelty effect of new technology may, in turn, bring about the learners' greater

motivation on specific learning behaviors. Such uncertainty over advanced technology was also reflected from the low-level circular continuity in VR co-creation; VR co-creators depended on lower-level chained behaviors to exchange information and reach community consensus to overcome the task complexity and to lower the sense of anxiety.

On the other hand, the ANCOVA results revealed the significant influences of the various co-creation environments on motivation. Further ANCOVA results displayed the significant effects of the different co-creation environments on intrinsic and extrinsic motivation. More specifically, on either intrinsic or extrinsic motivation, the influence of 3D VR co-creation was greater than that of 2D digital co-creation, and both were greater than that of paper-based co-creation as a means of structure visualization.

The results of the sequential behavioral analysis and ANCOVA supported what GmbH (2017), Hwang et al. (2017), Linnenbrink and Pintrich (2019), and Huang et al. (2020) claimed, namely that VR co-creation afforded contextualization, community, and relatedness, curiosity to stimulate make-believe fantasy, and learners' intrinsic and extrinsic motivation (Lepper et al., 2005; Lepper & Hodell, 1989).

There are further implications for future co-creation instruction and research. First, as reflected from the VR co-creators' persistent efforts on lower level knowledge co-construction behaviors, it is suggested that the instruction of VR co-creation be invested with even more class time to combat the inefficiency potentially existing in VR co-creation caused by higher technological complexity, as stated by Wang and Sun (2021), to allow abundant and mature higher level knowledge co-construction dialogues to occur. Secondly, appropriate preparatory learning activities, such as the pragmatic training on invitation strategies in virtual worlds, should also be further stressed as scaffolding mediators to elicit higher level knowledge co-construction behaviors and other expected learning outcomes to occur, especially when co-creation is imbued with a considerably high degree of equity and flexibility in leadership and task accountability for higher creativity (Lan, 2016).

There are two limitations to this study. First, due to the random sampling approach, the subgroup gender was not balanced. Specifically, the ratio of males to females in VR co-creation was 73:27, which was higher than that of the other two forms of co-creation, and therefore might to some extent reflect male students' higher motivation to interact with the 3D environment, as well as their strong eagerness for cognitive dissonance resolution and persistent self-looped behaviors to gain familiarity with new technology (Yeh et al., 2018). The second limitation lies in the lack of observation of the visualized interrelationship among the student co-creators. To gain an even deeper understanding of the social interpersonal structure embedded in the knowledge co-construction process of co-creation, it is suggested that social network analysis (SNA) be employed in future research. In terms of networks and graph theory, we may further visualize social relationships among co-creators for the advanced qualitative assessment of interpersonal co-creation activities.

Acknowledgement

This research was supported by the Higher Education Sprout Project of National Yang Ming Chiao Tung University (NYCU) and the Ministry of Education (MOE), Taiwan, as well as the Ministry of Science and Technology in Taiwan through Grant numbers MOST 110-2511-H-A49-009-MY2, MOST 107-2628-H-009-004-MY3, and MOST 105-2511-S-009-013-MY5. We would like to thank NYCU's ILTM (Interactive Learning Technology and Motivation, see: http://ILTM.lab.nycu.edu.tw) lab members and the students for helping conduct the experiment. We would also like to thank Pin-chia Feng for her support for and suggestions regarding the study as well as the reviewers who provided valuable comments.

References

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An Introduction to sequential analysis*. Cambridge University Press.

Bertolini, M., Scali, F., Poletti, G., Guerreschi, A., Fontana, F., & Hohenstein, U. T. (2018). Virtual portable art: Un percorso virtuale per le pietre incise di riparo tagliente [Virtual Portable Art: A virtual path for the engraved stones of Riparo Tagliente]. *Sezione di Museologia Scientifica e Naturalistica, 13*, 120-122. http://dx.doi.org/10.15160/1824-2707/1534

Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. In F. Biocca & M. R. Levy (Eds.), *Communication in the Age of Virtual Reality* (Vol. 15, pp. 57-124). Lawrence Erlbaum Associates.

Blau, I., & Shamir-Inbal, T. (2017). Re-designed flipped learning model in an academic course: The Role of co-creation and co-regulation. *Computers & Education, 115*, 69-81. https://doi.org/10.1016/j.compedu.2017.07.014

Bovill, C. (2019). Co-creation in learning and teaching: The Case for a whole-class approach in higher education. *Higher Education, 79*(6), 1023-1037. https://doi.org/10.1007/s10734-019-00453-w

Bovill, C., Cook-Sather, A., & Felten, P. (2011). Students as co-creators of teaching approaches, course design, and curricula: Implications for academic developers. *International Journal for Academic Development, 16*(2), 133-145. https://doi.org/10.1080/1360144X.2011.568690

Brandt, E., Messeter, J., & Binder, T. (2008). Formatting design dialogues - Games and participation. *International Journal of CoCreation in Design and the Arts, 4*, 51-64. https://doi.org/10.1080/15710880801905724

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A Systematic review. *The Internet and Higher Education, 27*, 1-13. https://doi.org/10.1016/j.iheduc.2015.04.007

Camargo-Borges, C., & Rasera, E. F. (2013). Social constructionism in the context of organization development: Dialogue, imagination, and co-creation as resources of change. *SAGE Open, 3*(2), 1-7. https://doi.org/10.1177/2158244013487540

Castaneda, L., Cechony, A., Bautista, A., & Pacampara, M. (2017). *All-school aggregated findings 2016-2017-VR*. https://drive.google.com/file/d/1nHd0tXr1DhGx8So1_2-HbkP2cJA5S61c/view

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10-32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Deci, E. L., & Ryan, R. M. (1980). The Empirical exploration of intrinsic motivational processes. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology* (Vol. 13, pp. 39-80). Academic Press.

Elsharnouby, T. H. (2015). Student co-creation behavior in higher education: The role of satisfaction with the university experience. *Journal of Marketing for Higher Education, 25*(2), 238-262. https://doi.org/10.1080/08841241.2015.1059919

Freina, L., & Ott, M. (2015, April). A Literature review on immersive virtual reality in education: State of the art and perspectives [Paper presentation]. International Scientific Conference Elearning and Software for Education, Bucharest, Romania. https://www.itd.cnr.it/download/eLSE%202015%20Freina%20Ott%20Paper.pdf

George, D., & Mallery, M. (2003). *Using SPSS for Windows step by step: A Simple guide and reference*. Allyn and Bacon.

Gergen, K. J., & Gergen, M. (2004). *Social construction: Entering the dialogue*. Taos Institute Press.

Giner, G. R., & Peralt Rillo, A. (2016). Structural equation modeling of co-creation and its influence on the student's satisfaction and loyalty towards university. *Journal of Computational and Applied Mathematics, 291*, 257-263. https://doi.org/10.1016/j.cam.2015.02.044

GmbH, D. (2017). *CoSpaces EDU: Create virtually anything in the classroom*. https://cospaces.io/edu/

Grover, S., Pea, R., & Cooper, S. (2015). Designing for deeper learning in a blended computer science course for middle school students. *Computer Science Education, 25*(2), 199-237. https://doi.org/10.1080/08993408.2015.1033142

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. *Journal of Educational Computing Research, 17*(4). https://doi.org/10.2190/7MQV-X9UJ-C7Q3-NRAG

Hsiao, I. Y., Lan, Y. J., Kao, C. L., & Li, P. (2017). Visualization analytics for second language vocabulary learning in virtual worlds. *Educational Technology & Society, 20*(2), 161-175.

Huang, W., Roscoe, R. D., Johnson-Glenberg, M. C., & Craig, S. D. (2020). Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning, 37*(3), 745-758. https://doi.org/10.1111/jcal.12520

Hwang, G.-J., & Chen, C.-H. (2017). Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioral patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology, 48*(4), 950-971. https://doi.org/10.1111/bjet.12464

Jans, S. D., Geit, K. V., Cauberghe, V., Hudders, L., & Veirman, M. D. (2017). Using games to raise awareness: How to co-design serious mini-games? *Computers & Education, 110*(1), 77-87. https://doi.org/10.1016/j.compedu.2017.03.009

Jarvis, L. (2019). Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media. *International Journal of Performance Arts and Digital Media, 15*(2), 239-241. https://doi.org/10.1080/14794713.2019.1637100

Keller, J. M. (2010). Generating and sustaining attention forethought. In J. M. Keller (Ed.), *Motivational Design for Learning and Performance: The ARCS Model Approach* (pp. 75-96). Springer.

Koch, M., Luck, K. v., Schwarzer, J., & Draheim, S. (2018, June). The Novelty effect in large display deployments – Experiences and lessons-learned for evaluating prototypes. 16th European Conference on Computer-Supported Cooperative Work-Exploratory Papers, Nancy, France. https://dl.eusset.eu/handle/20.500.12015/3115

Krause, S. (2017). *Build your own battle field: Using technology to understand the American revolution*. https://blog.cospaces.io/build-your-own-battle-field-30eddd18097d

Krokos, E., Plaisant, C., & Varshney, A. (2018). Virtual memory palaces: Immersion aids recall. *Virtual Reality, 23*(1), 1-15. https://doi.org/10.1007/s10055-018-0346-3

Lamb, R. (2014). Examination of allostasis and online laboratory simulations in a middle school science classroom. *Computers in Human Behavior, 39*, 224-234. https://doi.org/10.1016/j.chb.2014.07.017

Lan, Y. J. (2013). The Effect of technology-supported co-sharing on L2 vocabulary strategy development. *Educational Technology & Society, 16*(4), 1-16.

Lan, Y. J. (2014). Does second life improve mandarin learning by overseas Chinese students? *Language Learning & Technology, 18*(2), 36-56.

Lan, Y. J. (2016). The essential components of game design in 3D virtual worlds: From a language learning perspective. In S. M., L. B., & C. M. (Eds.), *Learning, Design, and Technology* (pp. 1-18). Springer.

Lan, Y. J. (2020). Immersion, interaction and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology, 24*(1), 1-15. https://doi.org/10125/44704

Landis, J. R., & Koch, G. G. (1977). The Measurement of observer agreement for categorical data. *International Biometric Society, 33*(1), 159-174. https://doi.org/10.2307/2529310

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184-196. https://doi.org/10.1037/0022-0663.97.2.184

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184-196. https://doi.org/10.1037/0022-0663.97.2.184

Lin, H. C.-S., Yu, S.-J., Sun, J. C.-Y., & Jong, M. S. Y. (2019). Engaging university students in a library guide through wearable spherical video-based virtual reality: Effects on situational interest and cognitive load. *Interactive Learning Environments*, 1-16. https://doi.org/10.1080/10494820.2019.1624579

Linnenbrink, E. A., & Pintrich, P. R. (2019). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313-327. https://doi.org/10.1080/02796015.2002.12086158

Lubicz-Nawrocka, T. M. (2018). Students as partners in learning and teaching: The Benefits of co-creation of the curriculum. *International Journal for Students as Partners, 2*(1), 47-63. https://doi.org/10.15173/ijsap.v2i1.3207

Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., & Rothbaum, B. O. (2017). The Use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry, 25*(3), 103-113. https://dx.doi.org/10.1097%2FHRP.0000000000000138

Mayer, R. E. (1996). History of instructional psychology. In E. d. Corte & F. E. Weinert (Eds.), *International Encyclopedia of Developmental and Instructional Psychology* (pp. 29-33). Pergamon Press.

Moro, C., Stromberga, Z., Raikos, A., & Stirling, A. (2017). The Effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education, 10*(6), 549-559. https://doi.org/10.1002/ase.1696

Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of Psychology: Educational Psychology* (pp. 103-122). Wiley.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* National Center for Research to Improve Postsecondary Teaching and Learning.

Schlacht, I. L., Mastro, A. D., & Nazir, S. (2017). Virtual reality for safety, entertainment or education: The Mars mission test. In V. G. Duffy (Ed.), *Advances in Applied Digital Human Modeling and Simulation* (pp. 75-83). https://doi.org/10.1007/978-3-319-41627-4_8

Schunk, D., P. Pintrich, & J. Meece. (2008). *Motivation in education: Theory, research, and applications*. Englewood Cliffs.

Sun, J. C.-Y. (2014). Influence of polling technologies on student engagement: An Analysis of student motivation, academic performance, and brainwave data. *Computers & Education, 72*, 80-89. https://doi.org/10.1016/j.compedu.2013.10.010

Sun, J. C.-Y., & Hsieh, P.-H. (2018). Application of a gamified interactive response system to enhance the intrinsic and extrinsic motivation, student engagement, and attention of English learners. *Educational Technology & Society, 21*(3), 104-116.

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

Sun, J. C.-Y., Kuo, C.-Y., Hou, H.-T., & Lin, Y.-Y. (2017). Exploring learners' sequential behavioral patterns, flow experience, and learning performance in an anti-phishing educational game. *Educational Technology & Society, 20*(1), 45-60.

Sun, J. C.-Y., Lin, C.-T., & Chou, C. (2018b). Applying learning analytics to explore the effects of motivation on online students' reading behavioral patterns. *International Review of Research in Open and Distributed Learning, 19*(2). https://doi.org/10.19173/irrodl.v19i2.2853

Thornhill-Miller, B., & Dupont, J.-M. (2016). Virtual reality and the enhancement of creativity and innovation: Under recognized potential among converging technologies? *Journal of Cognitive Education and Psychology, 15*(1), 102-121. https://doi.org/10.1891/1945-8959.15.1.102

Wang, H.-Y., & Sun, J. C.-Y. (2021). Real-time virtual reality co-creation: Collective intelligence and consciousness for student engagement and focused attention within online communities. *Interactive Learning Environments*, 1-14. https://doi.org/10.1080/10494820.2021.1928711

Yeh, Y.-L., Lan, Y.-J., & Lin, Y.-L. R. (2018). Gender-related differences in collaborative learning in a 3D virtual reality environment by elementary school students. *Educational Technology & Society, 21*(4), 204-216.

Yilmaz, K., Gizem, F., & Keser, H. (2020). Vertical versus shared e-leadership approach in online project-based learning: A Comparison of self-regulated learning skills, motivation and group collaboration processes. *Journal of Computing in Higher Education, 32*(3), 628-654. https://doi.org/10.1007/s12528-020-09250-2

An Eye Tracking Based Investigation of Multimedia Learning Design in Science Education Textbooks

Tugba Altan[1*] and Kursat Cagiltay[2]
[1]Kahramanmaras Sutcu Imam University, Turkey // [2]Middle East Technical University, Turkey //
tugbaaltan@ksu.edu.tr // kursat@metu.edu.tr
[*]Corresponding author

(Submitted April 20, 2021; Revised August 25, 2021; Accepted September 13, 2021)

ABSTRACT: This study investigated the effects of multimedia learning and visual design in a 6th grade science textbook on students' studying processes. This was accomplished by using eye tracking technology and by applying multimedia learning design and visual design principles to science textbooks. Eye tracking based testing was employed to evaluate the effects of multimedia learning and visual design principles on students' studying process as they interacted with the revised textbook chapter. The results revealed that the revised cell biology chapter facilitated answering achievement test questions and helped attentional focus on relevant images, as well as more successful integration of text and images during students' studying processes. These research findings can be used in the design process to develop science textbooks based on learner needs. The research also provides guidelines for designing similar multimedia learning materials. Thus, the research results may contribute both theoretical and practical implications for the multimedia learning design of science textbooks.

Keywords: Multimedia learning design, Visual design, Eye tracking, Textbook

1. Introduction

Textbooks are common learning materials in schools all around the world. Content, pedagogy, and visual design are common criteria categories used for textbook evaluation in the literature (Lemmer et al., 2008; Maun, 2006; Vinisha & Ramadas, 2013). There are three usual types of textbook evaluation studies including (1) survey method, (2) textbook analysis, and (3) experimental evaluation (Mikk, 2002). Qualitative research methods were also used to examine the quality of textbooks (Maun, 2006; Lemmer et al., 2008). All the evaluation methods for textbooks are helpful in investigating the instructional value of a textbook. However, studies generally do not examine students' learning experience directly while they are interacting with textbooks.

The visuals, and the integration of text and visuals in science textbooks were examined in textbook analysis studies. These studies focused on the criteria such as the type of visuals used, the proximity of text and visuals, and the integration of text and visuals (Slough et al., 2010; Vinisha & Ramadas, 2013). The integration of text and pictures and the visual design of the textbook can affect learning (Mayer, 2014; McCrudden & Rapp, 2015) so principles including Cognitive Theory of Multimedia Learning (CTML), Cognitive Load Theory, message design and visual design can be useful in textbook evaluation. Nevertheless, there is a gap for examining multimedia learning and textbook design from an instructional design perspective in the literature (Cheng et al., 2015; Mayer et al., 1996; Mayer et al., 1995; Peterson, 2016).

There are numerous studies related to multimedia learning and eye tracking that examined several principles such as multimedia, signaling, prior knowledge etc. of CTML for the integration of text and visuals in learning materials (Boucheix & Lowe, 2010; Jarodzka et al., 2010; Johnson & Mayer, 2012; Mason et al., 2013; Molina et al., 2018; Ozcelik et al., 2010; Schneider et al., 2018). However, there is limited research examining multimedia learning by using eye tracking with younger learners (Mason et al., 2013; Molina et al., 2018; Yang et al., 2018). Most studies were conducted with university level students rather than target learner groups and learning materials included very concise visuals in Mayer's studies as well as short texts in materials used for other studies (Jarodzka et al., 2017). Therefore, it is important to investigate different multimedia learning design principles at once in diverse materials with the target learner groups (Jarodzka et al., 2017). Moreover, there is still a paucity of research examining the multimedia learning design of textbooks, and existing research examines learning outcomes and applies experimental methods (Cheng et al., 2015; Mayer et al., 1996). Thus, there is a need in the literature to investigate multimedia learning textbook design with target level learners by applying several instructional design principles of the CTML at once.

This study investigated the evaluation process of a 6th grade science textbook used in public schools as a multimedia learning material. This was accomplished using eye tracking and by synthesizing the perspectives of

different participant groups (authors, graphical designer, students, and teachers). Multimedia learning design and visual design principles were also utilized to extend these perspectives based on the literature. The purposes of this study are to (1) examine multimedia learning design and the visual design of 6th grade science textbooks by using eye tracking methodology, and to (2) generate a guideline with several principles for evaluating the design of science textbooks.

2. Literature review

2.1. Textbook evaluation, multimedia learning, and visual design

In the literature, there are several studies evaluating textbooks through different research methods (e.g., Alpan, 2004; Karadağ et al., 2013; Keser, 2004; Lemmer et al., 2008; Maun, 2006; Mikk, 2002; Özay & Hasenekoğlu, 2007; Şahin, 2012; Uçar & Özerbaş, 2016; Vinisha & Ramadas, 2013). Survey studies are easy and valuable for revealing various criteria for judging the quality of a textbook from different sources such as authors, teachers, students, etc. However, different people value different criteria for evaluating textbooks (Mikk, 2002; Özdemir, 2007).

Textbook analyses examine various characteristics of textbooks based on strict criteria, using qualitative analytical methods such as content analysis, discourse analysis etc. (Abd-El-Khalicket al., 2017; Bierema et al., 2017; Mikk, 2002; Sharma & Buxton, 2015). Several characteristics of textbooks related to visual design including the type and the quality of the graphical or visual representations, the integration of visual representations with text (Gkitzia et al., 2011; Slough et al., 2010; Slough & McTigue, 2013; Vinisha & Ramadas, 2013), and changes in graphical representations in time (Lee, 2010) were explored by using textbook analysis. However, textbooks analysis has limitations such as difficulty in setting rules to decide all the important characteristics to be counted and the validity of collected data during the evaluation process (Mikk, 2002).

The literature shows that different research methods are helpful for examining different aspects of textbooks in evaluation processes. However, these methodologies do not examine the textbook design and development process and are mostly conducted after textbooks are published. Thus, there is a need to explore the textbook design and development process from an instructional design perspective. The criteria for textbook evaluation consist of various sub-criteria related to content, pedagogy, message design, visual design, and general features. For example, Young and Reigeluth (1988) emphasize the criteria for subject-matter content, social content, readability, and instructional design. Similar to general textbook evaluation, there are various criteria for evaluating science textbooks depending on different perspectives (Bösterli, 2015; Devetak & Vogrinc, 2013; Khine, 2013). For example, Bösterli (2015) developed a checklist including two different types of criteria/standard categories: set weighted standards and variable weighted standards. Previously, Devetak and Vogrinc (2013) defined three main criteria categories: general, textual, and pictorial.

The textbook analysis studies examining the visuals and the integration of text and visuals in science textbooks focused on the criteria such as the type of visuals used, the proximity of text and visuals, and the integration of text and visuals (Slough et al., 2010; Vinisha & Ramadas, 2013). However, none of the studies examined all the principles of visual design or message design in science textbooks using a holistic approach. The results of previous research showed that unrelated or decorative visuals and weak text-visual integration are common issues with science textbooks (Lee, 2010; Slough et al., 2010; Vinisha & Ramadas, 2013), and extraneous representations in textbooks remain an issue despite a widespread emphasis on using visuals which represent the learning content in the literature (Lee, 2010; Mayer, 2014).

Aside from research on visuals and the visual design of textbooks, there is a gap for examining multimedia learning textbook design from an instructional design perspective (Cheng et al., 2015; Mayer et al., 1996; Mayer et al., 1995; Peterson, 2016). A generative theory of textbook design was introduced supporting the use of annotated illustrations in science texts in the literature. The annotated illustrations exemplify the signaling principle of CTML, helping students' selection process of relevant words and relevant images. The results of other studies revealed that utilizing multimedia learning design in textbooks improves learning outcomes (Cheng et al., 2015; Mayer et al., 1996; Peterson, 2016).

The results of previous studies indicate that all the evaluation methods and criteria for textbooks and printed materials are helpful in investigating the instructional value of textbooks or other materials. However, there is an emphasis on a few common aspects of visual design and multimedia learning including the type of visuals used, the proximity of text and visuals, and the integration of text and visuals in the evaluation criteria (Cheng et al.,

2015; Lee, 2010; Mayer et al., 1996; Mayer et al., 1995; Peterson, 2016; Slough et al., 2010; Slough & McTigue, 2013; Vinisha & Ramadas, 2013). A limited amount of research examines the visual design of textbooks holistically (Alpan, 2004). None of the other studies examined multimedia learning and visual design at once, holistically, nor did they investigate all the appropriate principles of multimedia learning and visual design for science textbooks separately (Cheng et al., 2015; Lee, 2010; Mayer et al., 1996; Mayer et al., 1995; Peterson, 2016; Slough et al., 2010; Slough & McTigue, 2013; Vinisha & Ramadas, 2013). Furthermore, the studies did not investigate students' learning or studying experience directly while they are studying with the textbooks. Thus, there is a need in the literature to investigate the design and development process of science textbooks holistically, utilizing multimedia learning and visual design principles.

2.2. Eye tracking, multimedia learning design and visual design

Eye tracking technology is a unique way to evaluate design principles in the CTML. There are numerous studies related to multimedia learning and eye tracking in the literature that examined several principles such as multimedia, signaling, prior knowledge etc. of the CTML for the integration of text and visuals in learning materials (Boucheix & Lowe, 2010; Jarodzka et al., 2010; Johnson & Mayer, 2012; Mason et al., 2013; Molina et al., 2018; Ozcelik et al., 2010; Schneider et al., 2018). Researchers investigated learning outcomes using several principles of multimedia learning such as signaling (e.g., Boucheix & Lowe, 2010; de Koning et al., 2010; Ozcelik et al.., 2010), prior knowledge (e.g., Canham & Hegarty, 2010; Jarodzka et al., 2010), the modality effect and the effect of pacing on learner's attention (e.g., Meyer et al., 2010; Schmidth-Weigand et al., 2010), color coding (e.g., Ozcelik, et al., 2009), and spatial contiguity (e.g., Johnson &Mayer, 2012). Review results from eye tracking studies in science learning suggest designing digital learning materials for science education based on multimedia learning design principles (Yang et al., 2018). Eye tracking methodology may provide educational researchers an opportunity to link learning outcomes and cognitive processes, and this methodology also helps to evaluate learning materials and review the design principles in learning environments. It is important to investigate different multimedia learning design principles at once in diverse materials with the target learner groups (Jarodzka et al., 2017).

3. Methods

This study employed a design-based research (DBR) model to examine multimedia learning and the visual design of a 6th grade science textbook through eye tracking. DBR was used in this study because designing multimedia learning material requires iterative analysis, design, development, and redesign (McKenney & Reeves, 2013).

3.1. The science textbook

The literature suggests that multimedia learning design promotes learning when applied to science textbook design (Cheng et al., 2015; Mayer et al., 1996). Of particular interest are research on issues related to students' understanding of cell biology (Vijapurkar et al., 2014). Consequently, a cell biology chapter in a 6th grade science textbook was selected to investigate this issue further. In the previous phase of DBR, the multimedia learning and visual design of a cell biology chapter was evaluated and several issues with the chapter and students' needs were exposed. The issues and students' needs were irrelevant pictures, unclear directions, inappropriate use of color, number of visuals, alignment issues, image size, unclear images, inconsistent design, quality of paper/printing, text design, complex pictures, image quality, inappropriate use of image shape, imbalanced design, not motivating, inappropriate for learner level, misconception, inappropriate for individual differences, disconnected from daily life. In this phase, the issues were solved, and the cell biology chapter was revised using several principles of multimedia learning design (multimedia, coherence, segmenting, signaling, personalization, pre-training, and spatial contiguity), visual design (accurate design, alignment, aesthetic proportion-balance, clarity, color, consistency, harmony, image, simplicity, unity), and emotional design.

3.2. Participants

The selection criteria were that participants (1) should be 5th grade middle school students (2) should have little knowledge/understanding of the topic selected, (3) should show good performance in science class, (4) should not wear eyeglasses. Eighty students were selected as participants from one middle school based on their school

cumulatives and science grades. It was also critical to select students known to be good in science and successful at school because their reading comprehension levels were assumed to be higher than other students, allowing them to learn a new science topic and allowing researchers to control the difference in reading comprehension levels in different groups of learners. There is a correlation between reading comprehension and science achievement in the literature (Cromley et al., 2010). Participants should also have little knowledge/understanding of cell biology, since their learning experience was examined. Eye tracking sessions were conducted with 60 participants (26 boys, 34 girls) because they were available to participate the study after dealing with several issues such as parental consent etc. There were two groups and participants were assigned to the groups by using simple random grouping method. Thirty students participated in one group, while another thirty students participated in the other group. Their ages were between 9 and 12 years.

3.3. Eye tracking setup

A Tobii X2-60 mobile eye-tracker and Tobii mobile eye-tracker stand were used to conduct the eye tracking test in this study (see Figure 1). Earlier to the session, a demographic survey and a ten-minute prior knowledge test including seven questions about cell biology basics were administrated to each participant. Subsequently, each participant underwent a brief orientation session during the eye tracking session. Participants could have a "practice" session before the "study" eye tracking session began. The eye tracking session began immediately after calibration of the eye-tracker. Participants answered seven achievement test questions identical to those in the prior knowledge test.

Figure 1. Eye tracking setup

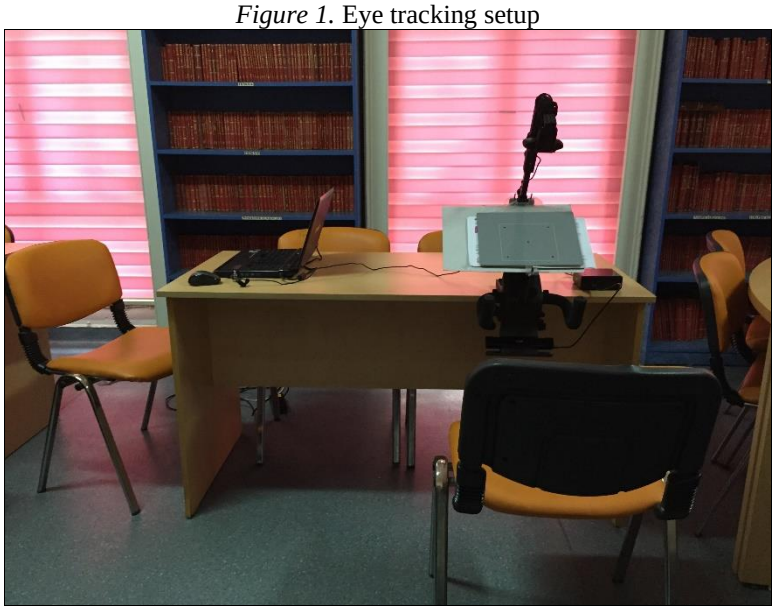
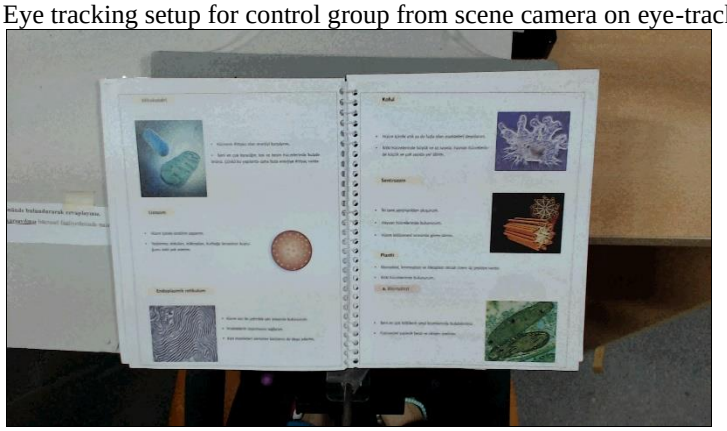


Figure 2. Eye tracking setup for control group from scene camera on eye-tracker stand

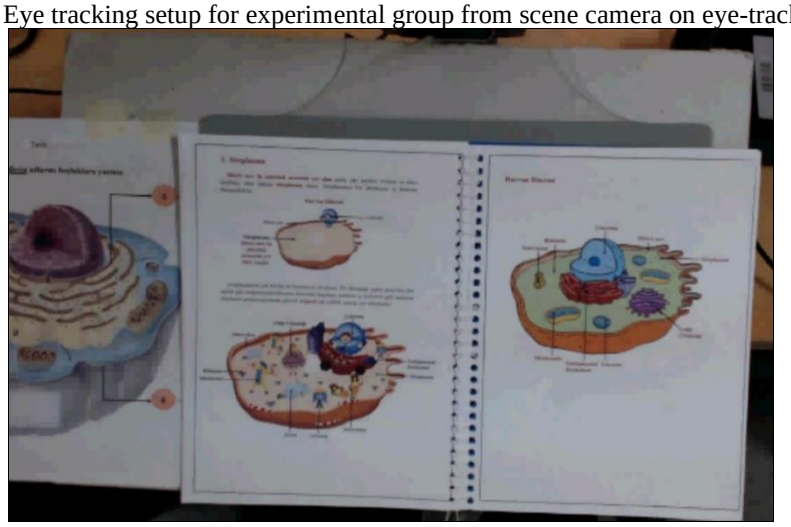


Participants studied for the achievement test questions by reading the textbook individually for an average of 40 minutes. First, they studied cell biology topics by themselves while their eye-movements were recording. Then, each question on a separate sheet of paper was presented and the researcher asked participants to answer these

questions one at a time. Participants tried to complete the achievement test using a printed copy of the cell biology chapter.

Participants in one group used the unrevised version of the cell biology chapter in the textbook (see Figure 2) and the second group used the revised version of the same chapter (see Figure 3).

Figure 3. Eye tracking setup for experimental group from scene camera on eye-tracker stand



3.4. Data sources and analysis

Eye movements and a voice recording of the eye tracking sessions, as well as the achievement test (r = .92), the prior knowledge test, and the student demographic survey were among the data sources for the study. The prior knowledge test was evaluated by experts for validity check. Areas of Interest (AOI) were created to calculate eye tracking data on relevant texts and relevant images in previously created scenes. Eye tracking measures including fixation duration, total fixation duration, proportion of fixations, visit duration, and visit count for relevant images and relevant texts were calculated.

Beside those eye tracking measures, transitions (integrative transitions, text-to-diagram transitions, and corresponding transitions) were counted in the eye tracking data. Achievement test performance and eye tracking data were analyzed using descriptive statistics and one-way ANOVA.

Additionally, the researcher used a standard test protocol during eye-tracking sessions and followed the protocol step by step for both groups.

4. Results

4.1. Achievement test results

Results revealed that both groups were similar to each other in terms of grades in science class, $F(1, 58) = .428$, $p = .51$ and in school as measured cumulatively $F(1, 58) = 3.97$, $p = .05$. Prior knowledge results revealed that the participants did not have much knowledge on cell biology topics, although some of them (3 participants in the control group and seven of them in the experimental group) already knew the definition of “cell” and some basic concepts.

A normality assumption was tested, and the results satisfied the assumption for the unrevised chapter group, but the results seemed not to satisfy the normality assumption for the revised chapter group. (Das & Imon, 2016; Field, 2016; Thode, 2002). However, ANOVA is robust to violations of the normality assumption, so it was used to compare means for achievement test scores in the two groups (Field, 2016). For the homogeneity of variance assumption, the results of Levene’s test showed that the assumption was not satisfied, $F(1, 58) = 29.98$, $p = .00$. But the sample sizes for both groups were equal and p value was decreased from .05 to .01 to conduct one-way ANOVA for these measures, as well as conducting Welch and Brown-Forsythe corrections ($p = .00$) (Field, 2016).

Table 1. One-Way ANOVA summary table for the effects of multimedia learning design on achievement test scores

Achievement test scores	*df*	*SS*	*MS*	*F*	*p*	η^2
Between-group	1	5096.81	5096.81	293.73	.00	.83
Within-group	58	1006.41	17.35			
Total	59	6103.23				

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

ANOVA results (see Table 1) revealed that there is a statistically significant difference in means for achievement test scores ($F(1, 59) = 293.73$, $p < .01$, $\eta^2 = .83$) between the revised chapter group ($M = 47.48$, $SD = 2.13$) and the unrevised chapter group ($M = 29.05$, $SD = 5.48$). The results indicated that participants' achievement test scores in the revised chapter group were better than in the unrevised chapter group, and the effect size was large.

4.2. Eye tracking results

A Mann-Whitney U non-parametric test was conducted to compare the means for the percentages of transitions in the unrevised and revised chapter groups, since the sample size is small for this eye tracking measure. The results showed that there was significant difference in means for integrative transitions ($U = 23.00$, $p < .05$), text-to diagram transitions ($U = 23.00$, $p < .05$), corresponding transitions ($U = .00$, $p < .05$) and proportion of corresponding transitions ($U = 23.00$, $p < .05$) between groups as summarized in Table 2. The result for the percentage of integrative transitions may imply that the participants in the revised chapter group ($M = 23.22$, $SD = 5.14$) attempted to integrate relevant words and relevant pictures significantly more than participants in the unrevised chapter group ($M = 15.55$, $SD = 4.34$). Similarly, the results for text-to-diagram transitions may imply that the participants in the revised chapter group ($M = 11.85$, $SD = 3.10$) attempted to integrate relevant words and relevant pictures significantly more than participants in the unrevised chapter group ($M = 7.74$, $SD = 2.16$).

Table 2. Mann-Whitney U test results table for the effects of multimedia learning design and visual design on transitions (in percentage)

Groups	*N*	Mean Rank	*U*	*Z*	*p*
Integrative transitions					
Unrevised chapter	15	9.53	23.00	-3.09	.00*
Revised chapter	11	18.91			
Text-to Diagram Transitions					
Unrevised chapter	15	9.53	23.00	-3.09	.00*
Revised chapter	11	18.91			
Corresponding transitions (total)					
Unrevised chapter	15	8.00	.00	-4.28	.00*
Revised chapter	11	21.00			
Proportion of corresponding transitions (Total)					
Unrevised chapter	15	8.00	.00	-4.28	.00*
Revised chapter	11	21.00			
Corresponding transitions (Text-to-diagram)					
Unrevised chapter	13	7.00	.00	-4.24	.00*
Revised chapter	12	19.50			
Proportion of corresponding transitions (Text-to-diagram)					
Unrevised chapter	13	7.08	1.00	-4.19	.00*
Revised chapter	12	19.42			

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

Another similar result was that the percentage of total corresponding transitions had a significantly higher mean ($U = .00$, $p < .05$) for the revised chapter group ($M = 18.87$, $SD = 3.58$) than for the unrevised chapter group ($M = 8.65$, $SD = 2.57$). The percentage of text-to-diagram corresponding transitions had a significantly higher mean ($U = .00$, $p < .05$) for the revised chapter group ($M = 9.58$, $SD = 2.15$) than for the unrevised chapter group ($M = 3.29$, $SD = 1.00$). The mean for proportion of total corresponding transitions was also significantly higher ($U = 1.00$, $p < .05$) for the revised chapter group ($M = .82$, $SD = .08$) than for the unrevised chapter group ($M = .49$, $SD = .08$). These results implied that the revised textbook chapter group was significantly more successful at integrating words and pictures than was the unrevised chapter group.

A normality assumption for ANOVA was satisfied for the majority of eye tracking measures ($p > .05$) except visit duration on relevant images in the revised chapter group, but ANOVA was conducted since this statistic is robust to violations of the normality assumption (Field, 2016). The homogeneity of variance assumption was satisfied for the visit count on relevant texts ($p > .05$) and the visit count on relevant images ($p > .05$). However, the data for total fixation duration on relevant texts ($p < .05$), total fixation duration on relevant images ($p < .05$), visit duration on relevant images ($p < .05$) and visit duration on relevant texts ($p < .05$) violated the assumption. Similarly, the homogeneity of variance assumption was violated for the proportion of fixations on relevant texts ($p < .05$) and the proportion of fixations on relevant images ($p < .05$) (Field, 2016). Although the assumption was violated for these measures, the sample sizes for both groups were equal and p value was decreased from .05 to .01 to conduct one-way ANOVA for these measures, as well as conducting Welch and Brown-Forsythe corrections ($p = .00$) (Field, 2016).

Table 3. Descriptive statistics for eye tracking measures on relevant images

Eye tracking measures		Unrevised chapter		Revised chapter	
	n	*M*	*SD*	*M*	*SD*
Proportion of fixations	30	.19	.05	.33	.08
Total fixation duration	30	6.19	2.78	10.52	4.06
Visit duration	30	9.37	3.14	15.18	4.86
Visit count	30	14.06	4.88	19.76	5.68

ANOVA results for eye tracking measures (see Table 3 and Table 4) revealed that means for the proportion of fixations on relevant images in the control group ($M = .19$, $SD = .05$) and the experimental group ($M = .33$, $SD = .08$) differed significantly from each other, $F(1, 58) = .60.86$, $p < .01$, $\eta^2 = .51$. For total fixation duration on relevant images, the difference in means between the experimental group ($M = 10.52$, $SD = 4.06$) and the control group ($M = 6.19$, $SD = 2.78$) was statistically significant, $F(1, 58) = 23.27$, $p < .01$, $\eta^2 = .29$.

Table 4. One-Way ANOVA summary table for the effects of multimedia learning design and visual design on eye tracking measures on relevant images

Variable and source	*df*	*SS*	*MS*	*F*	*p*	η^2
Proportion of fixations						
Between-group	1	.30	.30	60.86	.00*	.51
Within-group	58	.28	.00			
Total	59	.58				
Total fixation duration						
Between-group	1	281.72	281.72	23.27	.00*	.29
Within-group	58	702.24	12.11			
Total	59	983.96				
Visit duration						
Between-group	1	505.67	505.67	30.21	.00*	.34
Within-group	58	970.69	16.74			
Total	59	1476.37				
Visit count						
Between-group	1	487.24	487.24	17.39	.00**	.23
Within-group	58	1625.21	28.02			
Total	59	2112.45				

Note. $^*p < .01$, $^{**}p < .05$, η^2 = effect size.

Similarly, ANOVA was significant for visit duration on relevant images, $F(1, 58) = 30.21$, $p < .01$, $\eta^2 = .34$. The results showed a statistically significant difference in means between the experimental group ($M = 19.76$, $SD = 5.68$) and the control group ($M = 14.06$, $SD = 4.88$) for visit count on relevant images, $F(1, 58) = 17.39$, $p < .05$, $\eta^2 = .23$. All the results showed large effect sizes for the eye tracking measures. The results indicated that the participants in the experimental group showed significantly more interest in relevant images in the revised chapter than the control group showed in relevant images in the unrevised chapter. The experimental group also focused on relevant images significantly more often than did the control group.

The results showed that the means differed significantly for the proportion of fixations on relevant texts as summarized in Table 6, $F(1, 58) = .60.86$, $p < .01$, $\eta^2 = .51$. The results implied that the participants in the control group ($M = .81$, $SD = .05$) paid significantly more attention to relevant texts than the experimental group ($M = .67$, $SD = .08$) (see Table 5).

Table 5. Descriptive statistics for eye tracking measures on relevant texts

Eye tracking measures		Unrevised chapter		Revised chapter	
	n	*M*	*SD*	*M*	*SD*
Proportion of fixations (%)	30	.81	.05	.67	.08
Total fixation duration (%)	30	27.07	10.56	21.71	7.27
Visit duration (%)	30	44.73	10.14	32.91	6.56
Visit count (%)	30	18.30	5.82	24.68	5.81

The ANOVA result was significant for total fixation duration on relevant texts, F (1, 58) = 5.26, $p < .01$, $\eta^2 = .08$ (see Table 6). The result implied that the participants in the unrevised chapter group (M = 27.07, SD = 10.56) paid significantly more attention to relevant texts than in the revised chapter group (M = 21.71, SD = 7.27; Johnson & Mayer, 2012).

Table 6. One-Way ANOVA summary table for the effects of multimedia learning design and visual design on eye tracking measures on relevant texts

Variable and source	*df*	*SS*	*MS*	*F*	*p*	η^2
Proportion of fixations						
Between-group	1	.30	.30	60.86	.00*	.51
Within-group	58	.28	.00			
Total	59	.58				
Total fixation duration						
Between-group	1	432.16	432.16	5.26	.03*	.08
Within-group	58	4769.08	82.23			
Total	59	5201.24				
Visit duration						
Between-group	1	2094.36	2094.36	28.73	.00*	.50
Within-group	58	4228.39	72.90			
Total	59	6322.75				
Visit count						
Between-group	1	609.86	609.86	18.04	.00**	.24
Within-group	58	1960.58	33.80			
Total	59	2570.44				

Note. $^*p < .01$, $^{**}p < .05$, η^2 = effect size.

For visit duration, the means also differed significantly between the unrevised chapter and the revised chapter groups, F(1, 58) = 28.73, $p < .01$, $\eta^2 = .50$ (see Table 6). The result for visit duration on relevant texts implied that the participants in the unrevised chapter group (M = 44.73, SD = 10.14) spent much more time on relevant texts to get general information in order to complete the achievement test than the revised chapter group (M = 32.91, SD = 6.56) (Holmqvist et. al, 2011). ANOVA results were also significant for visit count on relevant texts, F(1, 58) = 18.04, $p < .05$, $\eta^2 = .24$ (see Table 6). The percentage of visit counts on relevant text in the revised chapter group (M = 24.68, SD = 5.81) were significantly higher than the old chapter group's percentage (M = 18.30, SD = 5.82). This result indicated that both groups may have showed different interest in relevant texts during eye tracking sessions.

5. Discussion

Research results showed that participants using the revised textbook chapter performed better on the achievement test than participants using the unrevised textbook chapter during eye tracking sessions, and the effect size was large. This result was similar to several research results suggesting that using several principles for multimedia learning and visual design facilitated learning (e.g., Alpan, 2004; Boucheix & Lowe, 2010; Cheng et al., 2015; Eitel et al., 2013; Johnson & Mayer, 2012; Mayer et al., 1996; Molina et al., 2018; Ozcelik et al., 2009; Ozcelik et al., 2010; Peterson, 2016). For example, Alpan's (2004) research results suggested that students using a booklet with holistic visual design scored significantly better than the other group. Boucheix and Lowe (2010) reported that color cues helped participants get higher scores than arrow cues or no cues in their study. Similarly, another study revealed that signals fostered learning and participants who studied with signaled material performed better than participants who studied the material with no signals (Ozcelik et al., 2010). Another research result showed that participants' retention and transfer performance increased when color

codes were used (Ozcelik et al., 2009). For the spatial contiguity principle, Johnson and Mayer's (2012) research results revealed that text and image integration helped participants in experiment groups to outperform other groups on transfer tests.

For the multimedia principle, one study showed that the group who received the summary with text and visuals performed as well as or better on recall questions and transfer problems than students in other groups (Mayer et al., 1996). In another study, learning outcomes were better based on recall and comprehension scores when visuals were presented in a self-paced format, both before and concurrently with the text rather than presenting only the text (Eitel et al., 2013). Molina et al. (2018) showed that presenting image and text together resulted in significantly higher post-test scores than presenting text only. Another study revealed that a modified textbook which aligned with several principles of multimedia learning (the multimedia principle, modality, the split attention effect, and the avoiding redundancy effect) helped students to perform better on conceptual knowledge, transfer and retention tests than students in a group using the standard textbook (Cheng et al., 2015). However, current research results regarding achievement test scores were different from some other research results in the literature because the post-test scores did not differ significantly between groups when using several multimedia principles (signaling, contiguity, and coherence) together, or using spot-light cues, or using proximity in multimedia learning materials in those studies (Clinton et al., 2016; de Koning et al., 2010; Molina et al., 2018).

Eye tracking results supported the impact of multimedia learning design and visual design on achievement test scores, as the participants in the revised chapter group attempted to integrate relevant text and relevant pictures significantly more often and were significantly more successful than participants in the unrevised chapter group. The research results were similar to Johnson and Mayer's (2012) research results. The participants in the revised chapter group paid significantly more attention to relevant images than the participants in the unrevised chapter group based on proportion of fixations, total fixation duration and visit duration. These results were similar to various studies in the literature (e.g., Boucheix & Lowe, 2010; de Koning et al., 2010; Eitel et al., 2013; Molina et al., 2018; Ozcelik et al., 2009; Ozcelik et al., 2010). There was a large effect size for each of the eye tracking metrics in this study.

Research results regarding proportion of fixations, total fixation duration and visit duration were different from some other research results in the literature (Clinton et al., 2016; Johnson & Mayer, 2012). Clinton et al. (2016) showed that participants with low prior knowledge had higher total fixation duration on relevant images in the control group than the experimental group. However, the participants in the revised chapter group paid significantly more attention to relevant images than the participants in the unrevised chapter group based on proportion of fixations, total fixation duration and visit duration in this study. Another study revealed that the proportion of fixations, total fixation duration and visit duration did not differ significantly between two groups in three different experiments using the spatial contiguity principle (Johnson & Mayer, 2012). The difference in current research results may arise from investigating several principles of multimedia learning and visual design holistically in this study. The student participants of the study were also between Piaget's concrete operational stage and formal operational stage, and this characteristic may be considered as a confounding variable in future studies (Piaget, 1997).

Research results also showed that participants in the revised chapter group were significantly more interested in relevant images than participants in the unrevised chapter group, based on visit count similar to other measures for cognitive processing (Holmqvist et al., 2011). This result aligned with the results related to transitions measured in this study. From an ease-of-use perspective taken from usability studies, higher visit counts show the difficulty of an instrument (Holmqvist et al., 2011), but this eye tracking measure on relevant images in textbooks and similar educational materials can be explored in future studies for similarity or differences in research results.

Eye tracking results for relevant text were different as participants in the unrevised chapter group paid significantly more attention to relevant text than did participants in the revised chapter group based on proportion of fixations, total fixation duration, and visit duration measures. The research result was similar to other research results indicating that using several multimedia design principles decreases the time spent on relevant texts for experimental groups (Eitel et al., 2013; Molina et al., 2018). Nevertheless, research result was different from some other research results in the literature (Clinton et al., 2016; Johnson & Mayer, 2012; Ozcelik et al., 2009). However, Ozcelik's et al. (2009) research results suggested that higher fixation duration on relevant texts results in worse performance. Thus, current research results may also suggest that participants may have spent more time on relevant texts when they didn't find the answer for the questions related to relevant images through their cognitive processes. Differently for relevant texts, the results indicated that the revised chapter group showed more interest in relevant texts than the unrevised chapter group based on visit count measure. Relevant texts in the revised chapter may provide more semantic informativeness than relevant texts in the unrevised chapter

(Holmqvist et al., 2011). The effect sizes were large. This result also aligned with the results related to transitions measured in this study.

This study also differed from other studies in the literature by testing participants' science achievement during the studying process from a usability testing perspective rather than using the post-test at the end of the studying/eye tracking sessions (e.g., Alpan, 2004; Boucheix & Lowe, 2010; Cheng et al., 2015; Clinton et al., 2016; de Koning et al., 2010; Eitel et al., 2013; Johnson & Mayer, 2012; Mayer et al., 1996; Mayer et al., 1995; Molina et al., 2018; Ozcelik et al., 2009; Ozcelik et al., 2010; Peterson, 2016).

In this study, the student participants' science achievement and cumulative school achievement were higher than the other students in the school. It was assumed that their reading comprehension was also better than the others, as there is a correlation between reading comprehension and science achievement in the literature (Cromley et al., 2010). However, one study revealed a negative correlation between reading comprehension ability and the first-pass fixation time on the text (Mason et al., 2013). For this reason, the reading comprehension characteristics of participants can be taken into account in future studies. This study was also conducted using a printed multimedia learning material, while other studies used digital versions of the multimedia learning materials, and student-reading behavior may differ when studying with printed learning material as compared to studying with digital learning material (Wallis, 2017).

In this study, all the results for eye tracking measures have both similarities and differences with other research results in the literature (e.g., Boucheix & Lowe, 2010; Clinton et al., 2016; de Koning et al., 2010; Eitel et al., 2013; Johnson & Mayer, 2012; Molina et al., 2018; Ozcelik et al., 2009; Ozcelik et al., 2010). The other studies examined only one principle or a few principles of multimedia learning design, but this study examined several multimedia learning design principles and visual design principles at once, and consequently, this approach may have provided different results from the literature. Therefore, examining multimedia learning design and visual design in a holistic approach can contribute to the literature by testing the design principles of science textbooks in real world settings, as investigating different multimedia learning design principles at once in diverse materials was suggested in the literature (Clinton et al., 2016; Jarodzka et al., 2017). The results can also contribute to the literature by examining multimedia learning design and visual design of science textbooks with young learners as the target learner group (Jarodzka et al., 2017; Mason et al., 2013; Molina et al., 2018).

6. Conclusions

This study investigated students' studying process using a 6th grade public school science textbook as a multimedia learning material. Research results revealed that multimedia learning and visual design of the revised cell biology chapter facilitated students' studying processes and helped attentional focus on relevant images and integration of relevant texts and relevant images during their learning processes when the cell biology chapter of a 6th grade science textbook was redesigned based on certain principles of CTML and visual design. Two categories were established: visual design and multimedia learning design.

The research results also supported using multimedia learning design and visual design holistically in science textbook design. Applying design principles holistically in the revised cell biology chapter resulted in significantly less time spent on relevant texts, significantly more successful integration of relevant texts and relevant images as well as significantly higher achievement scores in the revised chapter group as compared to unrevised old chapter group. Thus, design principles can be used in a holistic way in science textbook design, and certain visual design principles can assist designers in a detailed way when applying multimedia learning design principles to science textbooks. These research findings can be used by experts in the design process to develop a science textbook based on learner needs as well as by teachers to select appropriate science textbooks for students. The research also provides theoretical implications for investigating CTML for younger learners in further research.

6.1. Implications for practice

The results also provide exemplary principles for designing similar multimedia learning materials using a holistic approach. These principles can also be used as criteria for evaluating science textbooks. Principles of multimedia, coherence, segmenting, signaling, personalization, pre-training, spatial contiguity, and split attention were among the selected CTML principles that can be used in science textbook design.

The principles of visual design resulting from this study consisted of 20 principles addressing visuals and overall layout (e.g., clarity, aesthetic proportion-balance, alignment, consistency, and unity) in science textbooks. The visual design principles used in the revision process included clarity, image, alignment, consistency, quality of paper-printing, simplicity, using blank space effectively, color selection, unity, accurate design, aesthetic proportion-balance, and design variations. Learner level, motivational elements using emotional design, individual differences, and establishing connection with daily life by using the cell factory analogy were also addressed during the revision process. Emotional design was applied to the revised chapter by using bright or vivid colors, especially red to guide learners' attention in overall design including titles and cues. Other colors in illustrations or images were selected among bright colors for the same purpose. Another revision related to emotional design included using comic characters for cell biology analogies to appeal to students as suggested in the literature (Mayer & Estrella, 2014; Stark et al., 2018).

6.2. Implications for future research

This study investigated student studying process using a science textbook that is used in Turkish public schools as a multimedia learning material. Future research may investigate the student learning process while using textbooks as multimedia learning materials in other fields such as math, language, social studies etc. as well as investigating learning process with a science textbook in other grade levels, and in different countries. Future studies can examine the difference between groups with first pass and second pass reading and inspection, and pattern analysis in eye-tracking results (Mason et al., 2013) can be applied. Cognitive load was not examined in this study, but future studies can explore whether there is a significant difference between the unrevised chapter group and the revised chapter group based on cognitive load for both relevant images and relevant texts. The research procedures can be replicated to investigate the multimedia learning design and visual design of digital materials for science education in future studies. The implications of the study may also be tested for digital science textbook design by using similar research procedures.

Acknowledgement

The authors are grateful to Professor Richard E. Mayer for his help and advice during the design process of the revised science textbook chapter.

References

Abd-El-Khalick, F., Myers, J. Y., Summers, R., Brunner, J., Waight, N., Wahbeh, N., Zeineddin, A. A., & Belarmino, J. (2017). A Longitudinal analysis of the extent and manner of representations of nature of science in U.S. high school biology and physics textbooks. *Journal of Research in Science Teaching, 54*(1), 82-120. https://doi.org/10.1002/tea.21339

Alpan, G. B. (2004). Effect of the graphic in textbooks on the achievement and attitudes towards the course of the students. *Educational Sciences and Practice, 3*(6), 193-209. http://ebuline.com/pdfs/6sayi/ebu_6_4.pdf

Bierema, A. M.-K., Schwartz, R. S., & Gill, S. A. (2017). To what extent does current scientific research and textbook content align? A Methodology and case study. *Journal of Research in Science Teaching, 54* (8), 1097-1118. https://doi.org/10.1002/tea.21399

Boucheix, J.-M., & Lowe, R. K. (2010). An Eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction, 20*(2), 123-135. https://doi.org/10.1016/j.learninstruc.2009.02.015

Bösterli, B. K. (2015). Checklist for competence-oriented textbooks in science. *American Journal of Educational Research, 3*(11), 1450-1454. https:// doi: 10.12691/education-3-11-16

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*(2), 155-166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Cheng, M., Chou, P., Wang, Y., & Lin, C. (2015). Learning effects of a science textbook designed with adapted cognitive process principles on grade 5 students. *International Journal of Science and Mathematics Education, 13*, 467-488. https://doi.org/10.1007/s10763-013-9471-3

Clinton, V., Cooper, J. L., Michaelis, J., Alibali, M. W., & Nathan, M. J. (2016). Revising visuals based on instructional design principles: Effects on cognitive load and learning. In C. Was, F. J. Sansosti, & B. J. Morris (Eds.), *Eye tracking technology applications in educational research* (pp. 195-218). IGI Global.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A Domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 102*(3), 687-700. https://doi.org/10.1037/a0019452

Das, K. R., & Imon, R. A. H. M. (2016). A Brief review of tests for normality. *American Journal of Theoretical and Applied Statistics, 5*(1), 5-12. https://doi.org/10.11648/j.ajtas.20160501.12

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from complex animation: seeing is understanding? *Learning and Instruction, 20*(2), 111-122. https://doi.org/10.1016/j.learninstruc.2009.02.010

Devetak, I., & Vogrinc, J. (2013). The Criteria for evaluating the quality of the science textbooks. In M. S. Khine (Ed.), *Critical analysis of science textbooks: Evaluating instructional effectiveness* (pp. 3-16). Springer. https://doi.org/10.1007/978-94-007-4168-3_1

Eitel, A., Scheiter K., & Schüler, E. (2013). How inspecting a picture affects processing of text in multimedia learning. *Applied Cognitive Psychology, 27*, 451-461. https://doi.org/10.1002/acp.2922

Field, A. (2016). *An Adventure in statistics: The Reality enigma*. SAGE https://edge.sagepub.com/field-adventures-in-stats2/student-resources/chapter-1/action-plan

Gkitzia, V., Salta, K., & Tzougraki, C. (2011). Development and application of suitable criteria for the evaluation of chemical representations in school textbooks. *Chemistry Education Research and Practice, 12*(1), 5-14. https://doi.org/10.1039/C1RP90003J

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A Comprehensive guide to methods and measures*. Oxford University Press.

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research, 10*(1), 1-18. https://doi.org/10.16910/jemr.10.1.3

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146-154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Johnson, C. I., & Mayer, R. E. (2012). An Eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*(2), 178-191. https://doi.org/10.1037/a0026923

Karadağ, M., Dülgeroğlu, İ., & Ünsal, Y. (2013). The Evaluation of secondary education 9th grade physics textbook in accordance with the views of teachers. *Gazi University Journal of Gazi Educational Faculty, 33*(3), 549-568. https://app.trdizin.gov.tr/makale/TVRneE1ESTJOZz09/ortaogretim-9-sinif-fizik-ders-kitabinin-ogretmen-gorusleri-dogrultusunda-degerlendirilmesi-

Keser, H. (2004). Evaluation of the computer course textbooks used in the 4th grade of primary school in terms of design principles. *The Journal of Turkish Educational Sciences, 2*(3), 261-280. https://app.trdizin.gov.tr/makale/TkRrek9EWTI/ilkogretim-4-sinif-bilgisayar-ders-kitaplarinin-gorsel-tasarim-ilkelerine-gore-degerlendirilmesi

Khine, M. S. (2013). Analysis of science textbooks for instructional effectiveness. In M. S. Khine (Ed.), *Critical Analysis of Science Textbooks: Evaluating instructional effectiveness* (pp. 303-310). Springer. https://doi.org/10.1007/978-94-007-4168-3_15

Lee, V. R. (2010). Adaptations and continuities in the use and design of visual representations in us middle school science textbooks. *International Journal of Science Education, 32*(8), 1099-1126. https://doi.org/10.1080/09500690903253916

Lemmer, M., Edwards, J., & Sello, R. (2008). Educators' selection and evaluation of natural sciences textbooks. *South African Journal of Education, 28*, 175–187. https://eric.ed.gov/?id=EJ1149996

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education, 60*(1), 95-109. https://doi.org/10.1016/j.compedu.2012.07.011

Maun, I. (2006). Penetrating the surface: The Impact of visual format on readers' affective responses to authentic foreign language texts. *Language Awareness, 15*(2), 110-127. https://doi.org/10.1080/09658410608668854

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43-71). Cambridge University Press.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*(1), 64-73. https://doi.org/10.1037/0022-0663.88.1.64

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33*, 12-18. https://doi.org/10.1016/j.learninstruc.2014.02.004

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). Generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43*(1) (1995), 31-43. https://doi.org/10.1007/BF02300480

McCrudden, M. T., & Rapp, D. N. (2015). How visual displays affect cognitive processing. *Educational Psychology Review, 29*(3), 1-17. https://doi.org/10.1007/s10648-015-9342-2

McKenney, S., & Reeves, T. C. (2013). Educational design research. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications technology* (pp. 131-140). Springer. https://doi.org/10.1007/978-1-4614-3185-5

Meyer, K., Rasch, T., & Schnotz, W. (2010). Effects of animation's speed of presentation on perceptual processing and learning. *Learning and Instruction, 20*(2), 136-145. https://doi.org/10.1016/j.learninstruc.2009.02.016

Mikk, J. (2002). Experimental evaluation of textbooks and multimedia. In S. Selander & M. Tholey (Eds.), *New educational media and textbooks, The 2nd IARTEM Volume* (pp. 121 - 140). Stockholm Institute of Education Press. https://files.eric.ed.gov/fulltext/ED472706.pdf

Molina, A. I., Navarro, O., Ortega, M., & Lacruz, M. (2018). Evaluating multimedia learning materials in primary education using eye tracking. *Computer Standards & Interfaces, 59*, 45-60. https://doi.org/10.1016/j.csi.2018.02.004

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behaviour, 26*(1), 110-117. https://doi.org/10.1016/j.chb.2009.09.001

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An Eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53*(2), 445-453. https://doi.org/10.1016/j.compedu.2009.03.002

Özay, E., & Hasenekoğlu, İ. (2007). Lise–3 biyoloji ders kitaplarındaki görsel sunumda gözlemlenen bazı sorunlar [Some observed issues in visual presentation of high school 3rd grade biology textbooks]. *Journal of Turkish Science Education 4*(1), 80-91. https://asosindex.com.tr/index.jsp?modul=articles-page&journal-id=1795&article-id=301148

Peterson, M. O. (2016). Schemes for integrating text and image in the science textbook: Effects on comprehension and situational interest. *International Journal of Environmental & Science Education, 11*(6), 1365-1385. https://doi.org/10.12973/ijese.2016.352a

Piaget, J. (1997). *The Origin of intelligence in the child* (M. Cook, Trans.). Routledge. https://doi.org/10.1037/11494-000

Schmidth-Weigand, F., Kohert, A., & Glowalla, U. (2010). A Closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100-110. https://doi.org/10.1016/j.learninstruc.2009.02.011

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A Meta-analysis of signaling affects learning with media. *Educational Research Review, 23*, 1-24. https://doi.org/10.1016/j.edurev.2017.11.001

Sharma, A., & Buxton, C. A. (2015). Human–nature relationships in school science: A Critical discourse analysis of a middle-grade science textbook. *Science Education, 99*(2), 260-281. https://doi.org/10.1002/sce.21147

Slough, S. W., McTigue, E. M., Kim S., & Jennings, S. K. (2010). Science textbooks' use of graphical representation: A Descriptive analysis of four sixth grade science texts, reading. *Reading Psychology, 31*(3), 301-325. https://doi.org/10.1080/02702710903256502

Slough, S. W., & McTigue, E. (2013). Development of the graphical analysis protocol (GAP) for eliciting the graphical demands of science textbooks. In M. S. Khine (Ed.), *Critical Analysis of Science Textbooks: Evaluating instructional effectiveness* (pp. 17-30). Springer. https://doi.org/10.1007/978-94-007-4168-3_2

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A Mixed-methods study using eye tracking. *Computers & Education, 120*, 185-196. https://doi.org/10.1016/j.compedu.2018.02.003.

Şahin, M. (2012). The Evaluation of textbooks in terms of message design principles. A*hi Evran University Journal of Kırşehir Education Faculty (KEFAD/JKEF), 13*(3), 129-154. https://dergipark.org.tr/en/pub/kefad/issue/59486/854938

Thode, C. H., Jr. (2002). *Testing for normality*. Marcel Dekker, Inc.

Uçar, C., & Özerbaş, D. S. (2016). Evaluation of the secondary school 5th grade science textbooks according to visual design principles. *Kastamonu Education Journal, 25*(4), 1373-1388. https://dergipark.org.tr/tr/pub/kefdergi/issue/30766/332473

Vijapurkar, J., Kawalkar, A., & Nambiar, P. (2014). What do cells really look like? An Inquiry into students' difficulties in visualising a 3-D biological cell and lessons for pedagogy. *Research in Science Education, 44*(2), 307-333. https://doi.org/10.1007/s11165-013-9379-5

Vinisha, K., & Ramadas, J. (2013). Visual representations of the water cycle in science textbooks. *Contemporary Education Dialogue, 10*(1), 7-36. https://doi.org/10.1177/0973184912465157

Wallis, C. (2017). *A Textbook dilemma: Digital or paper?* https://hechingerreport.org/textbook-dilemma-digital-paper/

Yang, F.-Y., Tsai, M.-J., Chiou, G.-L., Lee, S. W.-Y., Chang, C.-C., & Chen, L.-L. (2018). Instructional suggestions supporting science learning in digital environments based on a review of eye tracking studies. *Educational Technology & Society*, *21*(2), 28–45.

Young, M. J., & Reigeluth, C. M. (1988). *Improving the textbook selection process, fastback 275*. Phi Delta Kappa Educational Foundation. https://eric.ed.gov/?id=ED299699

Online Community Building in Distance Education: The Case of Social Presence in the Blackboard Discussion Board versus Multimodal VoiceThread Interaction

Julian Chen and Tatiana Bogachenko*
Curtin University, Australia // julian.chen@curtin.edu.au // tetiana.bogachenko@curtin.edu.au
*Corresponding author

(Submitted July 2, 2021; Revised September 13, 2021; Accepted October 12, 2021)

ABSTRACT: Establishment of online communities in distance education has been linked to improved engagement, retention, and learning outcomes. This study investigates how online community building was fostered in the text-based Discussion Board (DB) and multimodal VoiceThread (VT) in one of the postgraduate units offered by Open Universities Australia. Specifically, it delves into how social presence – encompassing affective, cohesive, and interactive indicators in the Community of Inquiry (CoI) – was facilitated on both platforms. Findings show that VT multimodal postings triggered more instances of social presence than DB postings across all three indicators. VT communication also bolstered a continuous and inclusive discourse by bringing participants closer by addressing members' names and mentioning posts made by others. It is concluded that multimodality afforded by VT can be more advantageous for online collaboration and engagement. Suggestions for course design to establish stronger social presence and for evaluation of multimodal platforms are also offered.

Keywords: Distance education, Social presence, Online communities, VoiceThread, Multimodality

1. Introduction

Distance education, such as that provided by Open Universities Australia (OUA), plays an integral role in the tertiary sector. It offers a sustainable option that enables adult students to continue their studies remotely, which bears relevance to the global pandemic that has severely disrupted in-class schooling. Without distance education, students would have discontinued their studies. In our School of Education, academics teaching OUA units have endeavoured to encourage students to become more involved in the Blackboard Discussion Board (DB). The linear, text-based method of posting and commenting, however, seems to lack the dynamic and multimodal feel of social networking (e.g., Facebook or Instagram). Barren discussion forums, manifested in lacklustre and sporadic interactions, are not uncommon in OUA units. Although a sense of community is not a prerequisite for collaborative online learning (Lowenthal & Snelson, 2017), it has been found to foster students' online engagement and participation, resulting in better learning outcomes (e.g., Campbell & Mislevy, 2013; Lambert & Fisher, 2013; Liu et al., 2009; Sadera et al., 2009).

These pedagogical concerns motivated us to experiment with VoiceThread (VT), an asynchronous multimodal platform, in one of our OUA units. VT offers multimodality that affords users to choose their favourable mode(s) of posting for communication such as audio, video, image, and PDF besides the standard text-based posts. We compared the quality and quantity of students' online discussions on VT and DB in order to ascertain whether multimodality can mitigate the impact of virtual distance (Watts, 2016). In this article, we report on how both platforms were utilised to facilitate *social presence* within the framework of Community of Inquiry (CoI; see Garrison et al., 1999). Specifically, we quantified Social Presence Density (SPD) in student postings to better capture the instances of social presence – "the ability of participants in the [CoI] to project their personal characteristics into the community, thereby presenting themselves to the other participants as 'real people' " (Garrison et al., 1999, p. 89). These findings render useful implications for research and teaching in distance education.

2. Literature review

2.1. Student retention and online community building

Studies on student participation in online courses show high attrition rates due to both student-related and institution-related factors (Boston et al., 2012; Burns, 2013; Campbell & Mislevy, 2013). Moore and Greenland (2017) found that OUA student dropouts resulted from personal factors such as anxiety, health, and family issues, as well as work overloads from job commitments. That said, the institution-related factors are associated with cultural, technological, and course content and design aspects such as difficulty with adjusting to a fully online learning environment (Rovai & Downey, 2010). Evidently, one of the most common issues in distance education is the sense of isolation to which it can lead (Burns, 2013; Khurana, 2016; Rovai & Downey, 2010; Willging & Johnson, 2009). For example, students in Burn's (2013) study conducted in a coaching program reported the feeling of loneliness and detachment in online learning. Willging and Johnson (2009) surveyed those online student dropouts and revealed that virtual isolation, resulting from a lack of interaction with peers and instructors, led to poor student retention.

Indeed, interaction is viewed as instrumental in the enhancement of online learning (Krista, 2011). It can make learning more personalised and meaningful, stimulate understanding through engagement and collaborative enquiry, and foster peer support and social networking (Anderson et al., 2005). Interaction in distance education can also be conceptualised as building "an online community"— defined as "a group of participants in a distance-based environment with a shared purpose and the relationship among them including their sense of belonging, trust, and interaction" (Sadera et al., 2009, p. 2). This mirrors the Vygotskian view of learning as a social process that promotes situated learning through co-construction of knowledge and scaffolding (Anderson et al., 2005). As such, online community building facilitates the development of critical thinking and communication skills (Liu et al., 2009). It also sustains engagement, boosts motivation, and creates a connection with online community members, thus leading to better learning outcomes and an overall satisfaction with the online course (Watts, 2016). Higher levels of engagement further increase the sense of belonging and help students remain engaged throughout the course (Campbell & Mislevy, 2013). Therefore, it is crucial to investigate how technological advancements could facilitate such development of online communities in distance education. To achieve this goal, we ground our study in the CoI framework with a focus on one of its key elements, *social presence*, as it allows us to identify the nuanced indicators of social presence in online communication (Lowenthal & Dunlap, 2020).

2.2. Community of Inquiry (CoI) as a theoretical framework

The seminal CoI framework proposed by Garrison et al. (1999) guides the development and evaluation of online courses, operationalised within three key elements: *social presence*, *cognitive presence*, and *teaching presence*. Social presence comprises affective aspects and personalisation of online learning, while cognitive presence entails the ability of online learners to participate in meaning-making through communication. Teaching presence represents the design and facilitation of an online course. Taking Dewey's view of education as the process of discovery, the framework conceptualises learning as a social process rather than basic knowledge transmission (Lipman, 2003). It also assists in shortening the "transactional distance" that is typically experienced by online learners, thus alleviating potential misunderstandings that may be exacerbated by psychological and communication gaps in online learning (Moore, 1993 as cited in Marmon, 2018; see also Watts, 2016). Supporting critical thinking and co-construction of knowledge in turn leads to effective learning (Lipman, 2003). These three CoI elements were found to correlate with student learning outcomes and retention rates, thereby making it a viable framework for online course design and evaluation (Meyer, 2013).

Since its inception, the CoI has been widely adopted and adapted by course developers and researchers (see, for example, Borup et al., 2012; Hughes et al., 2007; Khurana, 2016; Wu, 2015), and a call for more empirical studies on its application across online contexts has also been proposed in order to refine the framework (Lowenthal & Dunlap, 2020). Hence, our study aims to examine the development of online communities via multimodal online discussion tools while providing further evidence for the dynamics of social presence in this investigated phenomenon. To better situate this study in the current body of research, the construct of social presence is further discussed below.

2.2.1. Social presence

Despite inconclusive in the literature, definitions of social presence in distance learning have revolved around Short's et al. (1976) original notion that indicates the extent to which telecommunication can foster interaction among participants (as cited in Lowenthal & Snelson, 2017). Swan (2017) highlights that the definition of social presence has been further crystalised in recent CoI scholarship as: "the ability of participants to identify with a group, communicate openly in a trusting environment, and personal and affective relationships progressively by way of projecting their individual personalities" (Garrison, 2016, as cited in Swan, 2017, pp. 4-5). In line with the original CoI framework, this updated notion caters for engagement and motivation, feelings of immediacy or closeness, and awareness of other participants in the community (Rourke et al., 1999).

For distance learning, social presence is particularly vital as it infuses the human elements of face-to-face communication (e.g., body language) into online interaction, thus fostering a sense of belonging (Marmon, 2018). The nature of social presence (i.e., mutual participation, support through shared experiences, negotiation of meaning) is also conducive to cognitive presence realised in deeper learning (Lipman, 2003). Given its vibrancy, social presence has been tested out against different models to further verify this construct, such as Social Presence Model (Whiteside, 2007, 2015) and Social Connectedness Design framework (van Tryon & Bishop, 2009, as cited in Dikkers et al., 2017), and its suitability for data analysis. For instance, Gunawardena's et al. (2001) survey and focus group interviews with postgraduate students in the USA and Mexico revealed social presence to be one of the factors that either helped or hindered the development of online groups.

In our study, the CoI provides a viable framework to identify and evaluate the elements of social presence with their manifestations. As suggested by Garrison et al. (1999), three key categories encompass these elements: *emotional expression* (affective aspect), *open communication* (interactive aspect), and *group cohesion*. The first category incorporates word choices, emoticons, and symbols used to express emotions, humour, and self-disclosure. Things that fall into this category are commonly used for sharing real-life experiences, attitudes, and interests, paving the way for building a sense of belonging and trust, while reducing the feeling of isolation. The second category includes showing awareness of online activities and interacting with other participants via responses to posted messages, agreeing, complimenting, or expressing appreciation for discussion contributions. Finally, group cohesion is realised in mutual commitments to community-building through two-way personalised dialogues rather than one-directional communication. This includes using salutations and phatic expressions, addressing posting contributors by name, and using inclusive pronouns to refer to the group as a unity.

Hughes et al. (2007) adapted Rourke's (1999) CoI framework to examine social presence in text-based postings by three interprofessional groups of students at a university. Lowenthal and Dunlap (2020) demonstrated how modification of Rourke's (1999) framework could translate into the comparison of social presence in group discussions. Similarly, Swan and Shih (2005) used Rourke's framework to calculate the *social presence density* of text-based online discussion platforms. While CoI was initially geared towards text-based communication in the late 20th century (Garrison et al., 1999), the uptake of multimedia in the digital era has allowed for multimodal communication. Hence, it is pertinent to compare different modes of instruction regarding their capacities to foster social presence (Lambert & Fisher, 2013).

King (2008) was among the first to compare social presence in text-based and voice-based asynchronous communication. Following Rourke et al. (1999), she calculated Social Presence Density (SPD) as the number of social presence indicators divided by the number of words in messages. In her study, text-based postings triggered higher affective and communicative elements of social presence, whereas voice-based ones contained slightly more elements of cohesive interaction. She also found that when given a choice, students preferred to use text for posting. Wu (2015) compared SPD in both text and voice messages used by Chinese university students. He found higher SPD in voice communication regarding affective and interactive components, but text communication showed higher cohesion. While this finding seems to contradict those of King (2008), voice communication was arranged synchronously in this case, and the addition of some spoken elements, such as laughter or intonation, was also coded. Khurana (2016) took another approach to assessing social presence elements triggered by digital platforms in course modules, namely, text-based forums, VoiceThread, Voki, and Vocaroo. Different from how SPD was computed previously, Khurana (2016) calculated SPD by dividing the number of social presence instances by the number of posts (not words). All three elements of social presence showed higher density in the multimedia discussions compared to the text-based forums.

Besides different course designs and approaches to identifying SPD in prior studies, the multimodal aspect of online discussions has been less explored in finding interpretations, as has been the impact on results of policies of obligatory postings for unit completion. For example, voice postings also include suprasegmental ways of expressing emotions such as pitch and intonation. In King's (2008) study, however, no adjustment was made to

code different aspects of modality such as nonverbal features. Wu (2015) also acknowledged that many instances of suprasegmental expressions might have been omitted or not specified in his coding scheme given the cumbersome nature of this task (cf. Rourke et al., 2001). Finally, students in Khurana's (2016) study were expected to use a different mode (e.g., Voki) to make at least one post per module, although this "requirement" imposed on them also compromised the uninhibited nature of students' posting behaviour.

These observed issues highlight the primacy of coding subtle social presence elements that are easily eschewed or overlooked. Our study intends to address the aforementioned shortcomings and present a clearer picture of social presence that is mirrored in affective, interactive, and cohesive elements of VT and DB communications.

2.3. Discussion Board vs. VoiceThread

The Blackboard Discussion Board (DB, henceforth) is a widely utilised online discussion space for asynchronous communication (Kay, 2006). It provides a structured forum for students to read and comment on discussion topics, responding to each other's postings and archiving posts in a linear manner. Although DB posts are mainly text-based, students can also attach a file, an image, or include a link in their post. DB allows students to post and process the content materials more deeply at their own pace before posting (e.g., Hrastinski, 2008). However, this asynchronous mode also poses challenges to online students. For example, it only generates delayed responses as opposed to its synchronous counterparts that allow for real-time communication (Kay, 2006). The linear interface requiring users to collapse longer threads and posts in order to view and comment is also time-consuming and cumbersome (Kay, 2006). Another downside of DB is that students can simply respond by recycling what has already been addressed in their peers' posts without adding their own perspectives to the discussion (Kirby & Hulan, 2016), especially when deliberately avoiding disagreements with peers (deNoyelles et al., 2014).

In response to these limitations, alternative online discussion platforms have been implemented in online learning. VoiceThread (VT, henceforth) is an example of such a platform that is seeing increasing use in higher education (e.g., Chan & Pallapu, 2012; Delmas, 2017; Donnelly at el., 2016; Fox, 2017; Hsu et al., 2014; Khurana, 2016). Different from DB in interface, VT collates multiple postings in the same space that are accentuated by multimodal commenting (text, audio, video, image, and PDF) without inundating users with linear text-based threads. VT's multimodality can enhance understanding in online communication (Delmas, 2017; Fox, 2017) because it allows users to "communicate emotion, personality, and other non-verbal cues conducive to better understanding and interpretation of meanings" (Ching & Hsu, 2013, p. 308), thus stimulating online interaction and participation (Delmas, 2017; Donnelly et al., 2016; Sato et al., 2017). For example, students in Fox's (2017) study found it helpful to see and/or hear the instructor and peers on VT owing to multimodal postings that make content processing more concrete and comprehensible. Similarly, Japanese language learners in Sato's et al. (2017) research were able to use paralinguistic cues, demonstrated by the Japanese instructor in video tutorials posted on VT, to help them deepen learning by imitating her expressions in intercultural communication.

Multimodality also lends itself to favourable conditions for online collaboration. For example, students in Ching and Hsu's (2013) study considered audio-commenting on VT more time-efficient in facilitating feedback exchange than text-typing. Hsieh (2012) found that VT enhanced Taiwanese university students' development of critical thinking and English presentation skills while promoting collaboration. Additionally, the multimodal posting options afforded by VT personalise communication (Delmas, 2017; Fox, 2017) ~~and~~ bridge the virtual distance between students and their lecturers (Donnelly et al., 2016). Hence, multimodal features can reinvigorate text-based conversation that tends to appear formal and one-dimensional, as is usually the case in DB. The ability to hear/see the instructor in audios/videos, the immediacy of support from the teacher, and the multimodal interactions with peers can help develop a personal connection and combat isolation (see Chen et al., 2020; Khurana, 2016; Sato et al., 2017).

Despite the positive claims made about VT in prior studies, challenges were also reported. For instance, nursing students in Fox's (2017) study initially felt nervous about recording themselves due to their accents and lack of confidence. However, their anxiety dissipated gradually and preferences for voice commenting over text commenting increased. Understanding how social presence is developed and supported by multimodal tools can better inform stakeholders such as institutions, course designers, teachers, and students of optimal solutions to address these challenges. Herein lies the purpose of our study that intends to address the main research question:

Can the use of VoiceThread enhance OUA students' online learning experience and engagement as manifested in the instances and density of social presence when compared to the Discussion Board?

3. Methodology

3.1. Setting and participants

This study was part of a larger project investigating the effectiveness of VT for asynchronous multimodal discussions versus DB discussions. One of the postgraduate OUA units offered by an Australian public university was chosen to trial VT for unit discussions. As the unit content covers new ways of utilising emerging technologies for distance education, it met our selection criterion. The unit was delivered in OUA Study Period 3 (SP3), 2018, for a duration of 11 weeks.

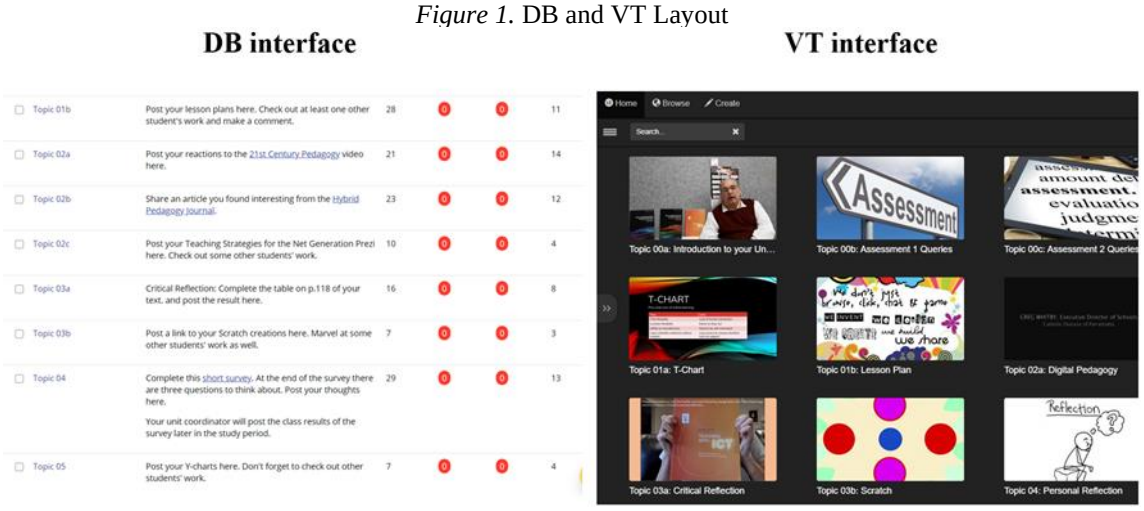
126 adult students were enrolled in the Master of Teaching Course via OUA and taking this unit in fulfillment of the Course. Since this unit has two tutorial groups in DB and VT, students were administratively allocated to one of the two discussion spaces through Blackboard. Despite the equal random assignment initially attempted, some students were late for enrolment in this unit, thus making it challenging to allocate equal numbers to both groups. As commonly observed in distance education, some students also withdrew from the unit at different times of SP3. Nevertheless, out of 75 students allocated to DB, 55 (73%) remained in the unit by the end of SP3. 51 students were assigned to VT, and 41 (80%) of them remained enrolled. Of the students who completed the unit, 24 (25%) were males, and 72 (75%) were females.

3.2. Data collection

Ethical clearance was sought and approved before the OUA unit commenced. Enrolled students were informed about the purpose of the study via Blackboard email announcement, which stated that participation in this project was entirely voluntary and no coercion or penalty would apply for early withdrawal. Students' real names were replaced with pseudonyms in the posting examples provided below.

Both VT and DB discussion spaces included separate sections for student introductions (week one) and questions about assignments, but only the 10 content sections (weeks two to eleven) were used for data analysis. In contrast with previous research, we kept topics and tasks the same for both text-based (DB) and multimodal (VT) platforms in the two piloted student groups. Each topic section contained a prompt posted by the instructor, such as a question, article, or video for the students to respond to, or instructions to create their own resources. On DB, the discussion prompt was linearly specified in the section topic, whereas instructions in the VT interface were included in the teacher's initial post within each section. Figure 1 illustrates the two different layouts.

Figure 1. DB and VT Layout



Students responded by creating a new thread or continuing the one initiated by the teacher or their peers. Participation was also managed differently: postings were not obligatory (though encouraged), nor were students marked for their online participation. This enabled us to ascertain how students and their teacher made personal

choices of using different modes of interaction, and how this influenced the patterns of communication on both platforms.

3.3. Data analysis

3.3.1. Coding

Student postings were extracted from both platforms and coded following the scheme developed by Rourke et al. (1999), while adjustments for multimodal platforms from Wu (2015) and Khurana (2016) were also considered. Hence, our coding scheme included three categories of social presence: *affective*, *interactive*, and *cohesive* (see detail in the Literature Review section) and their indicators as discussed above (see Table 1 for examples). Following Wu (2015), each indicator was given an abbreviation code to identify and tally the total number of instances throughout the dataset. To address non-verbal communications in audio postings, we adopted Khurana's (2016) coding scheme to capture vocal and non-verbal ways of expressing emotions through stress, intonation, and "extra sounds" (p. 64), such as giggling.

Table 1. Social presence categories (based on Wu's (2015) and Khurana's (2016) adaptation of Rourke's et al. (1999) codes, extended)

Categories	Indicators and codes	Examples
Affective	Expression of emotions (SP-AE)	• *Conventional expression of emotions:* I really enjoyed; (which worries me); it was so rewarding; I may be going mad....; oh my goodness; This video made me feel so so so so sad • *Unconventional expression of emotions:* :D ; … ; !!! ; FOR FREE!; haha; LOTS; If it hits the candle, it bursts! • *Non-verbal expression of emotions:* sigh, giggle, excessive gestures, and facial expressions
	Use of humour (SP-AU)	• I'm a bit late to the party here; Sorry for those who had to read this long post. I'm more of a traditionalist and still happy to type. Haha; To get my "total teacher geek" on,…
	Self-disclosure (SP-AS)	• For work I taught myself how to use R, a statistical and graphing software. • I've gone to numerous concerts small and large and I've seen the sound guys actually […] running the system off an iPad
Interactive	Continuing a thread (SP-IC)	• RE:
	Quoting from others' messages (SP-IQ)	• Not found
	Referring explicitly to others' messages (SP-IR)	• I really liked the comments you mentioned about parents and responsibility.
	Asking questions (SP-IA)	• Are you able to see it?
	Complimenting, expressing appreciation (SP-ICE)	• Great idea!; Thank you for your feedback.
	Expressing agreement (SP-IE)	• I agree with both of you; I agree with the other comments that
Cohesive	Vocatives (SP-CV)	• Mentioning names (e.g., as John said)
	Addresses or refers to the group using inclusive pronouns (SP-CA)	• It seems as though we all have quite similar experiences with our previous education in regards to exposure to technology.
	Phatic expressions, salutations (SP-CP)	• Hi all; Howdy; Afternoon; Hope this makes sense; Cheers

Borup et al. (2012) found that students appreciated "seeing" their lecturer when studying online as a means of simulating a face-to-face class and that facial expressions, posture, and gestures add to the sense of social

presence in distance learning. The students who authored video posts were articulate with their body language. Therefore, we carefully identified and coded the salient non-verbal cues used to display emotions in video postings, especially when gestures and facial expressions were shown more strongly than usual (e.g., eye-rolling combined with hand gestures) or were accompanied by intonational change. Table 1 presents our coding scheme, highlighted using colour-coding and illustrated by corresponding examples.

3.3.2. Unit of analysis and SPD

Coding was conducted by analysing posting content. While content analysis can utilise different units of categorisation, such as words, sentences, paragraphs, themes, or messages, Rourke et al. (2001) recommend using a single post as a unit of analysis for online communication. This method streamlines the coding procedure and avoids ambiguity since it becomes clearer to track where the post starts and ends. As Khurana (2016) asserts, focusing on a post as a unit of analysis is also convenient for multimedia messages. Indeed, multimodal communications on VT would make it hard to compare the length of posts. For example, some students tended to elaborate more in audio posting than in text posting. Speaking also involves different language structures and patterns (e.g., shorter and less complex) compared to written texts.

In contrast to a single sentence or paragraph, a post provides a context to which the codes can be assigned. For instance, Khurana (2016) explains how inclusive pronouns (we, us, our) are coded when they refer to the student group itself and left out when they refer to other groups or people in general. This guidance was helpful as we located similar examples in our data (e.g., "It actually reminded me of what *we* [SP-CA] read in the first chapter of the readings this week," referring to the group; "*We* would then have a quick review on what we learnt during the lesson," referring to this student as a teacher and his class). Other subtle discursive features were also noted and carefully coded. For example, "thank you" can be used to show appreciation of another person's post or just as a polite closure (i.e., phatic expression), and indirect questions were raised without question marks. Some indicators may be culturally specific, such as the use of humour.

Finally, SPD was calculated in order to compare the quantity and quality of the two discussion platforms (Khurana, 2016; Rourke et al., 2001). The use of a single post as a unit of analysis means that the SPD was calculated in relation to the number of posts rather than the number of words. To illustrate, an SPD of 3 means that on average, there are three instances of social presence in a post in a given selection (a thread, topic, or the whole discussion space). Figure 2 shows how several instances of the same type of social presence category can be identified within the same post. In other words, the total number of instances within the platform (as well as within each category) is then divided by the total number of posts by students (e.g., 352 instances in DB ÷ 139 posts in DB = 2.53 SPD for DB). Table 2 in the Findings section presents all the numbers and results for this study.

Figure 2. Example of post coding (colour added to indicate the categories, beginning, and ending of the coded text)

Hi Anne, [SP-CP; SP-CV]

I just had a look at your lesson plan [SP-AS]. Excellent, [SP-ICE]I would like to be in your class..... [SP-AE] wish my english classes were as interesting as that.. [SP-AS] Thanks for sharing. [SP-ICE]

Victoria. [SP-CP]

3.3.3. Intercoder reliability

To ensure the reliability of the results, the process of analysis involved multiple stages and coders. In the first stage, a combined 10 per cent of the VT and DB posts were individually coded by the research assistant (RA) and one of the project investigators, and their coding results were compared and discussed. Intercoder reliability was calculated based on the percentage of agreement (De Wever et al., 2006). Initially, a low level of coder agreement was reached at 0.62. After cross-examination and discussion on the discrepancies, an adjusted level of agreement was reached at 0.71, which is deemed as reliable due to the high level of subjectivity (Rourke et al., 2001). The RA recoded all the posts taking into account the discussed inconsistencies before another iteration of intercoder reliability was run with one of the authors, whereby it increased to 0.99. Four months later, the RA did

a final round of coding to ensure consistency in which instances were checked with the project lead when any doubts arose.

4. Findings

The results of the SPD calculations are presented in Table 2 and discussed in relation to social presence and online community-building. Key patterns, omissions, and additional observations are outlined and exemplified.

Table 2. Social Presence Density (SPD) for DB and VT respectively

Categories	Indicators and codes	DB[1]	VT[2]
Affective	• Expression of emotions (SP-AE)	53 (0.38)	229 (0.98)
	• Use of humor (SP-AU)	3 (0.02)	13 (0.05)
	• Self-disclosure (SP-AS)	76 (0.54)	121 (0.52)
	Total instances in Affective category	132	363
	*SPD for Affective category**	0.94	1.56
Interactive	• Continuing a thread (SP-IC)	46 (0.33)	207 (0.89)
	• Quoting from others' messages (SP-IQ)	0	0
	• Referring explicitly to others' messages (SP-IR)	2 (0.01)	71 (0.30)
	• Asking questions (SP-IA)	16 (0.11)	21 (0.09)
	• Complimenting, expressing appreciation (SP-ICE)	33 (0.23)	159 (0.68)
	• Expressing agreement (SP-IE)	17 (0.12)	43 (0.18)
	Total instances in Interactive category	114	501
	SPD for Interactive category	0.82	2.15
Cohesive	• Vocatives (SP-CV)	26 (0.18)	210 (0.90)
	• Addresses or refers to the group using inclusive pronouns (SP-CA)	7 (0.05)	8 (0.03)
	• Phatics, salutations (SP-CP)	73 (0.52)	163 (0.70)
	Total instances in Cohesive category	106	381
	SPD for Cohesive category	0.76	1.64
Total of SP instances		352	1245
# of posts by students in all topics		139	232
Overall SPD		2.53	5.36

Note. [1]Number of instances followed by SPD in brackets. [2] Number of instances followed by SPD in brackets.

Overall, SPD is more than twice as high in VT postings as on DB (5.36 vs. 2.53), and consistently higher on VT across all the three categories of social presence. If we focus on the Affective and Cohesive categories, both show higher density on VT and are almost twice as high as on DB (1.56 vs. 0.94; 1.64 vs. 0.76). The density discrepancy in the Interactive category is even more prominent in comparison (2.15 vs. 0.82). This consistent finding suggests that, given the same content and tasks, the multimodality afforded by VT means that the platform serves as a more viable venue for the development of social presence and online community.

A closer look at the Interactive category revealed that the VT forum hosted longer and more "continuous" discussions. Given the differences in the interface and structure of communication, it is not surprising that VT had fewer but longer threads (2.7 vs 8.3 posts per thread on DB and VT, respectively; see Chen et al., 2020). Within these longer threads, interactive social presence was facilitated on numerous occasions where students "explicitly referred to peer messages" – with an SPD indicator of 0.01 and 0.30 for DB and VT, respectively. Students also continued threads without referring directly to the content of their messages (e.g., complimenting others and expressing appreciation), reflecting a density that was three times higher on VT than DB. Expression of agreement, another way to engage and keep a conversation going, also exhibited a slightly higher density on VT than DB (0.18 vs. 0.12).

Interestingly, within the same category, the "questioning" indicator revealed a different pattern. Slightly more questioning posts were found on DB than VT, albeit insignificant in density (0.11 and 0.09). On both platforms, students posted questions asking for help with technology or resources, requesting feedback, and making attempts to clarify ideas of peers. This may indicate that questioning is seen as an integral part of interaction in distance education regardless of the medium of communication.

Within the Cohesive category, VT students were more inclined to address each other by name (0.90 vs. 0.18 on DB). This may also be attributed to the fact that VT involved longer threads with multiple participants within the

same thread. Hence, there was a need to specify to whom the message (response, question, comment) referred. Some also mentioned their own names in the posts (e.g., "Hi everyone, John here") – which was not discussed in prior studies but can be seen as contributing to the Cohesive element. Regarding the use of inclusive pronouns, DB students were slightly more likely to deploy them than their VT counterparts (0.05 vs. 0.03), indicating that communication on DB was geared more towards the group than individuals. One plausible explanation is that more students participated on DB than VT throughout SP3 (36/55=65% on DB vs. 21/41=51% on VT), and the VT group developed other forms of whole-group referencing in addition to inclusive pronouns. Other pronouns or nouns that indirectly refer to the group can also foster a sense of belonging (e.g., "fellow students," "all," "everyone").

Of the three indicators in the Affective category, only expression of emotions had a density that was almost three times higher on VT than DB (0.98 and 0.38, respectively). This is not surprising given that multimodal postings render more channels for expressing emotions through intonation and body language. That said, students in both groups used humour sparingly, resulting in similarly low-density indicators. The density of self-disclosure was also roughly at the same level. Similar to the employment of questions, this may indicate that students at postgraduate level tend to use examples and personal experiences to support their opinions (e.g., "When I compare my schooling exposure to the exposure children have now its is completely different." Note that No correction was made in grammatical errors/typos in students' original postings unless comprehension was hindered.). Both groups also shared their struggles with certain unit material or technological demands (e.g., "Im naturally not very tech savvy so this course is challenging me and teaching me alot :)")

A few additional features of these discussion spaces have been noted during data analysis. For example, it is worth mentioning that shorter posts were found to contain more instances of SP than longer ones on both platforms. This may be due to the fact that longer posts were primarily associated with the unit content material and responding to weekly tasks, whereas shorter ones functioned as expressing gratitude, emotions, questions or agreement. This is demonstrated in the following example:

DB: "This looks great Gregory! [SP-ICE; SP-CV; SP-AE] Is primary connections a text? [SP-IA]"
VT: "Hi Max [SP-CP; SP-CV], it could possibly be me - teething issues [SP-AU] haha [SP-AE]. I tried jpg and png. Then just tried to upload as word document... [SP-AS; SP-AE]"

Another interesting finding stems from the change in SPD trajectory throughout the unit— rather than a steady increase in SPD as the unit progressed, the density fell more in certain topic sections than in others, with a few spikes in density throughout the unit (see Appendix for the distribution of social presence instances across topics). This may suggest that the type of tasks, questions, or prompts may impact the effectiveness of online community-building. The highest SPDs were recorded across the same topics on both platforms: *2a Digital Pedagogy* (exchanging feedback on the video), *6a Cyber Safety* (reflecting on a video related to cyber safety), and *8b sharing thoughts about MOOCs and online communities*. These tasks encouraged students not only to respond to multimodal prompts with critical views but also to reflect on personal feelings and experiences, thus fostering social presence. Conversely, the two topics ending with the lowest SPD instructed students to only post links to external resources (e.g., "Post a link to your curated resources on cybersafety here. Check out other students' lists."). This suggests that task type and phrasing may influence the density of social presence, which leads to different levels of student task engagement.

In Topic 10 (the final topic), only VT students provided feedback on the unit since DB communication had sharply declined and almost no student posts were being made by week 10. Surprisingly, this final VT discussion triggered the highest level of SPD across all topics on both platforms. It included posts such as "Thanks for what has been a really enjoyable, and eye-opening unit" or "Thank you for your comments as well and … everybody else's that have posted." As VT was the only platform to feature posts in this topic, it is impossible to compare. However, the high level of social presence evidenced by the positive and appreciative nature of these posts implies the overall satisfaction, enjoyment, and positive learning experiences among OUA students in an online community hosted by VT.

Besides the calculated instances and observed patterns found in SPD, it is also instructive to discuss aspects that were absent from the posts. For instance, some posts (mostly on DB) displayed no phatic expressions and salutations at all (83/139 [0.59] on DB compared to 81/232 [0.34] on VT). It was also noted that some posts in both VT and DB discussions contained no text content. That is, some DB posts contained a link only with no explanation provided, whereas some VT posts included a visual only. However, there were no VT posts which exclusively contained a link (if so, all links were introduced or explained). Where an image or another form of media was posted, a follow-up post explaining the previous post would also be provided. VT also enables a

multimedia attachment (e.g., a chart or a picture) to be immediately displayed within the post without taking further steps by clicking to download and opening the file as in the DB interface.

Finally, we discovered certain aspects of social interaction in online community-building outside the scope of the existing social presence codes. For instance, students were keen on inviting peers or the instructor to respond to their initial posting or to offer suggestions. Such examples were particularly evident on VT, often featuring a concluding remark such as "Looking forward to your responses," "have a look and let me know what you think," or "I'd like to hear your thoughts and I look forward to seeing everyone else's T-Pack Y Chart as well." While these are coded as phatic expressions (SP-CP), they can also serve the additional social cohesive-interactive function of inviting others to further engage in the discussion, showing awareness of others' presence, acknowledging their (dis)agreement, and stimulating critical thinking. This may be considered as another indicator for inclusion in the task of refining the existing social presence framework.

5. Discussion and implications

Overall, findings in our study support claims made by prior research in that while both discussion platforms aim to support social presence in online community-building, VT's multimodality is more advantageous for online collaboration and engagement than text-based DB (Ching & Hsu, 2013; Delmas, 2017; Fox, 2017). This is evidenced by the increased interaction and personalisation in VT discussions as the unit progressed. Interestingly, the interactional aspect of social presence was not just more pronounced on VT but was situated in different patterns of communication. This was mainly due to the higher number of cohesive VT posts contributing to longer threads — students not only posted their individual responses to the task prompts but engaged further by using the space as a sounding board for their own ideas and providing input on the ideas of their peers via comments. The richer and deeper social presence promoted by multimodal discussion platforms such as VT can in turn facilitate a stronger online community. Multimodality can also enable personalisation in online discussion and community-building. Not only did VT provide multiple options for posting (thus accommodating individual differences and preferences), but the seamless unit design allowed students to flexibly choose any medium of communication without feeling obliged to pick one over another (cf. Khurana, 2016; King, 2008; Wu, 2015). As evidenced by our study, more tight-knit interpersonal connections were formed when multimedia options were provided, particularly through expressing emotions, addressing peer names, and referring to each other's messages. This also corroborates previous findings in that being able to use audio messages to hear each other's voices facilitates social presence as it develops a higher rapport and trust among the participants (Gunawardena et al., 2001; Khurana, 2016) and helps avoid misunderstanding (Aragon, 2003). Furthermore, students mentioning their own names in the posts also allowed for a greater degree of personalisation in posts which simulated 'real' conversation with peers or instructors.

Both platforms, however, seemed to have provided similar opportunities for self-disclosure (referring to one's personal experiences). Students in both groups equally shared their successes and struggles with the unit-related activities and content, as well as disclosing examples of challenges from their lives outside the unit. This may indicate that students were able to capitalise on real-life examples to support and strengthen their arguments. While making connections to their real-life experiences and teaching practices, it also made their responses more genuinely engaging. Above all, "putting yourself out there" in a public virtual space was no longer considered intimidating in an established online community.

Instances of asking questions were also equally found on both platforms, suggesting that students felt comfortable seeking help or feedback from other group members. Though not directly referring to the content of posts, requesting help and offering solutions helped continue the conversation. Within the existing CoI framework, this is an indicator of teaching presence (see Garrison et al., 1999). It is worth noting that the instructor was not the only one who provided answers as peer support was also found to be conducive to social presence, group interaction, and cohesion. Future research may seek to delve deeper into the effects of different online discussion platforms on the types of questioning, patterns of responses, and follow-up responses.

While overall SPD on VT remained higher than that of DB throughout the unit, SPD was also contingent upon the task type on both discussion platforms. Examination of social presence indicators across topic sections revealed that personalised discussion prompts spawned more student responses regardless of the medium. These tasks invited students to provide self-reflection, thoughts and feelings about an issue raised in the instructor's prompt. Though these prompts were mainly aimed at eliciting students' understanding of the tasks, they also provided avenues for students to develop solidarity by sharing genuine feelings about an issue uncovered or expressing empathy for others' challenging real-life experiences, including cyberbullying or struggles with

technological issues. This finding further supports Lowenthal and Dunlap's (2020) assertion that the nature of a task has an impact on social presence. In particular, they found that SPD was higher when online discussions had a clear focus, such as specific aspects of the assignments, and called for more research on task impact across instructional contexts.

Indeed, this salient finding yields pedagogical implications for course design. While incorporating collaborative learning activities to enhance social presence in online courses (Aragon, 2003) may not always be feasible, it shows that tasks designed to encourage learners to reflect on their own experiences can organically lead to heightened social presence. Consequently, a collegial online environment with high levels of social presence is "intrinsically valuable and educationally profitable" (Rourke et al., 1999, p. 60) as it creates a safe space for online students to openly share personal opinions and offer moral support.

Our findings also help refine the coding scheme originally developed by Rourke et al. (1999). For instance, Hughes et al. (2007) suggested distinguishing between explicit and implicit expression of emotions, re-categorising related indicators, and interpreting multiple meanings for words such as "Sorry." In our case, we also discovered different ways participants referred to a group that were not included in the original framework (such as "fellow students"). These instances flag up the inherent ambiguities of the original CoI framework, which in turn necessitates a concrete scheme of how coding can be done and what issues may arise in the process. Finally, while the CoI itself does not include practical steps to design online courses or improve social presence, case studies can be useful to test out this framework in that they lead to concrete implementation plans and evidence-based implications for distance education (Lowethal & Dunlap, 2020).

6. Limitations and conclusion

Social presence is a key element of the CoI framework that incorporates interactive, cohesive, and affective aspects of online communication during a course of study. The patterns identified in our research support the merit of the CoI framework, highlighting how particular indicators in social presence are manifested on different discussion platforms.

Course designers and online instructors may find the following implications germane to their settings: (1) longer and more continuous interaction patterns promoted by multimodal discussions foster social presence; (2) referencing each other's messages/names enhances a sense of belonging and group identity, thus contributing to the development of social presence; and (3) infusing "humanistic and empathetic" approaches in task design encourages more genuine sharing of feelings and real-life experiences – an integral part of online community-building. Above all, the refinement of the social presence coding scheme developed in our study could serve as a useful guide for future research into multimodal communication in distance education.

Despite its positive findings, this study also has some limitations. First, both groups were not equal in size despite our initial attempts to rectify this issue. As previously indicated, some students were late in enrolment, whereas some withdrew from the unit at different times of SP3, thus making an equal number unfeasible. Constraints posed by participant attrition are beyond researchers' control, however. Second, while higher SPD levels were found to be associated with certain task types, we did not track individual students' employment of social presence indicators. Individual differences and their impact on online community building could be another research area worth exploring. Other limitations are that only one OUA unit in a single Study Period was targeted and only student-student relationships and student posts were coded. Lecturer-student relationships and student interaction with the unit content (i.e., teaching presence and cognitive presence of the CoI framework) are also closely connected to social presence and deserve equal attention when discussion forums are compared. Since students used both platforms to seek help or feedback, it would also be helpful to investigate the question-response patterns in relation to the level of social presence and type of discussion space. These suggestions provide valuable insights and directions for further research.

Acknowledgement

We would like to thank both the Unit Coordinator, Dr. Martin Cooper, and the Learning Technologies Support Officer, Mr. Craig Sims, for their time and collaboration in this grant project (funded by Curtin's School of Education).

References

Anderson, T., Annand, D., & Wark, N. (2005). The Search for learning community in learner paced distance education: Or, 'Having your cake and eating it, too!' *Australasian Journal of Educational Technology, 21*(2), 222-241. https://doi.org/10.14742/ajet.1336

Aragon, S. (2003). Creating social presence in online environments. *New Directions for Adult and Continuing Education, 100*, 57–68. https://doi.org/10.1002/ace.119

Borup, J., West, R. E., & Graham, C. R. (2012). Improving online social presence through asynchronous video. *Internet and Higher Education, 15*, 195–203. https://doi.org/10.1016/j.iheduc.2011.11.001

Boston, W., Ice, P., & Burgess, M. (2012). Assessing student retention in online learning environments: A Longitudinal study. *Online Journal of Distance Learning Administration, 15*(2), 1-8. https://www.westga.edu/~distance/ojdla/summer152/boston_ice_burgess152.html

Burns, M. (2013). Staying or leaving? Designing for persistence in an online educator training programme in Indonesia. *Open Learning: The Journal of Open, Distance and eLearning, 28*(2), 141-152. https://doi.org/10.1080/02680513.2013.851023

Campbell, C. M., & Mislevy, J. L. (2013). Student perceptions matter: Early signs of undergraduate student retention/attrition. *Journal of College Student Retention, 14*(4), 467-493. https://doi.org/10.2190/CS.14.4.c

Chan, M., & Pallapu, P. (2012). An Exploratory study on the use of VoiceThread in a Business Policy course. *MERLOT Journal of Online Learning and Teaching, 8*(3), 223-237. https://jolt.merlot.org/vol8no3/chan_0912.pdf

Chen, J., Bogachenko, T., Sims, C., & Cooper, M. (2020). The use of VoiceThread as a multimodal digital platform to foster online students' task engagement, communication and online community building. In S. Dovchein (Ed.), *Digital communication, linguistic diversity and education* (pp. 175-201). Peter Lang. https://doi.org/10.3726/b15710

Ching, Y.-H., & Hsu, Y.-C. (2013). Collaborative learning using VoiceThread in an online graduate course. *Knowledge Management and E-Learning, 5*(3), 298–314. https://doi.org/10.34105/j.kmel.2013.05.021

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A Review. *Computers and Education, 46*(1), 6-28. https://doi.org/10.1016/j.compedu.2005.04.005

Delmas, P. M. (2017). Using VoiceThread to create community in online learning. *TechTrends, 61*, 595–602. https://doi.org/10.1007/s11528-017-0195-z

deNoyelles, A., Zydney, J. M., & Chen, B. (2014). Strategies for creating a Community of Inquiry through online asynchronous discussions. *MERLOT Journal of Online Learning and Teaching, 10*(1), 153-165. https://jolt.merlot.org/vol10no1/denoyelles_0314.pdf

Dikkers, A. G., Whiteside, A. L., & Tap, B. (2017). Understanding connections among definitions, theory, measurements, and practice. In A. L. Whiteside, A. G. Dikkers, & K. Swan (Eds.), *Social presence in online learning: Multiple perspectives on practice and research* (pp. 11-25). Stylus Publishing.

Donnelly, M. K., Kverno, K. S., Belcher, A. E., Ledebur, L. R., & Gerson, L. D. (2016). Applications of VoiceThread© technology in graduate nursing education. *Journal of Nursing Education, 55*(11), 655-658. https://doi.org/10.3928/01484834-20161011-09

Fox, O. H. (2017). Using VoiceThread to promote collaborative learning in on-line clinical nurse leader courses. *Journal of Professional Nursing, 33*(1), 20–26. https://doi.org/10.1016/j.profnurs.2016.08.009

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2-3), 87-105. https://doi.org/10.1016/S1096-7516(00)00016-6

Gunawardena, C. N., Nolla, A. C., Wilson, P. L., Lopez-Islas, J. R., Ramirez-Angel, N., & Megchun-Alpizar, R. M. (2001). A Cross-cultural study of group process and development in online conferences. *Distance Education, 22*(1), 85-121. https://doi.org/10.1080/0158791010220106

Hrastinski, S. (2008). A Study of asynchronous and synchronous e-learning methods discovered that each supports different purposes. *Educause Quarterly, 4*, 51–55. https://er.educause.edu/-/media/files/article-downloads/eqm0848.pdf

Hsieh, L. W. (2012). *Technology-supported project-based learning in a Taiwanese university oral communication course: A Case study* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Hsu, Y.-C., Ching, Y.-H., & Grabowski, B. L. (2014). Web 2.0 applications and practices for learning through collaboration. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and* technology (4th ed.) (pp. 747-758). Springer. http://dx.doi.org/10.1007/978-1-4614-3185-5_60

Hughes, M., Ventura, S., & Dando, M. (2007). Assessing social presence in online discussion groups: A Replication study. *Innovations in Education and Teaching International, 44* (1), 17–29. https://doi.org/10.1080/14703290601090366

Kay, R. H. (2006). Developing a comprehensive metric for assessing discussion board effectiveness. *British Journal of Educational Technology, 37*(5), 761–783. https://doi.org/10.1111/j.1467-8535.2006.00560.x

Khurana, C. (2016). *Exploring the role of multimedia in enhancing social presence in an asynchronous online course* (Unpublished doctoral dissertation). The State University of New Jersey, New Brunswick, NJ.

King, K. (2008). *Comparison of social presence in voice-based and text-based asynchronous computer conferencing* (Unpublished doctoral dissertation). Nova Southeastern University, Davie, FL.

Kirby, E. G., & Hulan, N. (2016). Student perceptions of self and community within an online environment: The Use of VoiceThread to foster community. *Journal of Teaching and Learning with Technology, 5*(1), 87-99. https://doi.org/10.14434/jotlt.v5n1.19411

Krista, J. L. (2011). *Digital communication applications in the online learning environment* (Unpublished doctoral dissertation). Oklahoma State University, Stillwater, OK.

Lambert, J. L., & Fisher, J. L. (2013). Community of Inquiry framework: Establishing community in an online course. *Journal of Interactive Online Learning, 12*(1), 1-16.

Lipman, M. (2003). *Thinking in education* (2nd ed). Cambridge University Press. https://doi.org/10.1017/CBO9780511840272

Liu, S. Y., Gomez, J., & Yen, C.-J. (2009). Community college online course retention and final grade: Predictability of social presence. *Journal of Interactive Online Learning, 8*(2), 165-182. https://digitalcommons.odu.edu/efl_fac_pubs/10/?utm_source=digitalcommons.odu.edu%2Fefl_fac_pubs%2F10&utm_medium=PDF&utm_campaign=PDFCoverPages

Lowenthal, P. R., & Dunlap, J. C. (2020). Social presence and online discussions: A Mixed method investigation. *Distance Education, 41*(4), 490-514. https://doi.org/10.1080/01587919.2020.1821603

Lowenthal, P. R., & Snelson, C. (2017). In search of a better understanding of social presence: An Investigation into how researchers define social presence. *Distance Education, 38*(2), 141-159. https://doi.org/10.1080/01587919.2017.1324727

Marmon, M. (2018). *Enhancing social presence in online learning environments*. IGI Global. https://doi.org/10.4018/978-1-5225-3229-3

Meyer, K. (2013). The case for the Community of Inquiry (CoI) influencing student retention. In Z. Akyol & D. R. Garrison (Eds.), *Educational communities of inquiry: Theoretical framework, research, and practice* (pp. 317-333). IGI Global. https://doi.org/10.4018/978-1-4666-2110-7.ch015

Moore, C., & Greenland, S. (2017). Employment-driven online student attrition and the assessment policy divide: An Australian open-access higher education perspective. *Journal of Open, Flexible and Distance Learning, 21*(1), 52-62. http://www.jofdl.nz/index.php/JOFDL/article/view/286/217

Rovai, A. P., & Downey, J. R. (2010). Why some distance education programs fail while others succeed in a global environment. *The Internet and Higher Education, 13*, 141–147. http://dx.doi.org/10.1016/j.iheduc.2009.07.001

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education, 14*(2), 50-71. https://www.learntechlib.org/p/92000/

Rourke, L., Anderson, T., Garrison, D., & Archer, W. (2001). Methodological issues in the content analysis of computer conference transcripts. *International Journal of Artificial Intelligence in Education,* 12(2001), 8-22.

Sadera, W. A., Robertson, J., Song, L., & Midon, M. N. (2009). The role of community in online learning success. *MERLOT Journal of Online Learning and Teaching, 5*(2), 277-284. https://jolt.merlot.org/vol5no2/sadera_0609.pdf

Sato, E., Chen, J. C., & Jourdain, S. (2017). Integrating digital technology in an intensive, fully online college course for Japanese beginning learners: A standards-based, performance-driven approach. *The Modern Language Journal, 101*(4), 756-775. https://doi.org/10.1111/modl.12432

Short, J., Williams, E., & Christie, B. (1976). *The Social psychology of telecommunications*. Wiley.

Swan, K. (2017). Multiple perspectives on social presence in online learning. In A. L. Whiteside, A. G. Dikkers, & K. Swan (Eds.), *Social presence in online learning: Multiple perspectives on practice and research* (pp. 3-10). Stylus Publishing.

Swan, K., & Shih, L. F. (2005). On the nature and development of social presence in online course discussions. *Journal of Asynchronous Learning Networks, 9*(3), 115–136. http://dx.doi.org/10.24059/olj.v9i3.1788

Watts, L. (2016). Synchronous and asynchronous communication in distance learning: A Review of the literature. *The Quarterly Review of Distance Education, 17*(1), 23–32. https://www.proquest.com/scholarly-journals/synchronous-asynchronous-communication-distance/docview/1794526758/se-2?accountid=10382

Willging, P. A., & Johnson, S. D. (2009). Factors that influence students' decision to drop out of online courses. *Journal of Asynchronous Learning Networks, 13*(3), 115–127. https://www.learntechlib.org/p/104037/

Wu, L. (2015). *Investigating social presence in the voice-based chat room and the text-based forum in the Chinese online learning context* (Unpublished doctoral dissertation). University of the West of England, Bristol, UK. http://eprints.uwe.ac.uk/27173

Appendix. Topical distributions of social presence instances and SPD

Tasks	DB instances / posts	DB density (SPD)	VT instances / posts	VT density (SPD)
1a Post your T-charts: Identifying the advantages and disadvantages of the technology you have chosen here. Check out some of your peer's T-charts.	43/24	1.79	124/29	4.27
1b Post your lesson plans here: Check out at least one other student's work and make a comment.	62/18	3.44	64/13	4.92
2a Post your reactions to the 21st Century Pedagogy video here.	55/14	3.92	98/15	6.53
2b Share an article you found interesting from the Hybrid Pedagogy Journal.	19/12	1.58	43/12	3.58
2c Post your Teaching Strategies for the Net Generation Prezi here: Check out some other students' work.	23/8	2.87	105/13	8.07
3a Critical Reflection: Complete the table on p.118 of your text. and post the result here.	18/9	2	64/13	4.92
3b Post a link to your Scratch creations here: Marvel at some other students' work as well.	15/4	3.75	110/17	6.47
4 Complete this short survey: At the end of the survey there are three questions to think about. Post your thoughts here. Your unit coordinator will post the class results of the survey later in the study period.	34/16	2.12	50/10	5
5 Post your Y-charts here: Don't forget to check out other students' work.	11/5	2.2	64/13	4.92
6a Watch this video then post your feelings and reflections about cybersafety.	33/9	3.66	130/17	7.64
6b Post a link to your curated resources on cybersafety here: Check out other students lists.	3/4	0.75	52/13	4
7a Post a link to a video or website you could use in teaching about cybersafety to a class: Click on some other students' links.	2/4	0.5	66/18	3.66
7b Post your PDF containing a Canva and lesson plan here: Enjoy viewing and commenting on other students' work.	3/2	1.5	108/20	5.4
8a Share an idea for a flipped or blended learning activity that uses technology here.	19/7	2.71	37/8	4.62
8b Share your thoughts on either the MOOC you joined or the online community you joined here.	12/3	4	27/4	6.75
9 Post your KWLH diagrams here.	0	0	56/12	4.66
10 If you wish you may post your constructive thoughts about this unit here and we will try to improve things based on your suggestions.	0	0	47/5	9.4

Guest Editorial: Creative Learning in Authentic Contexts with Advanced Educational Technologies

Rustam Shadiev[1], Wu-Yuin Hwang[2*] and Gheorghita Ghinea[3]
[1]Nanjing Normal University, Nanjing, Jiangsu Province, China // [2]National Central University, Jhongli, Taiwan, China // [3]Brunel University London, Uxbridge, Middlesex, United Kingdom // rustamsh@gmail.com // wyhwang@cc.ncu.edu.tw // george.ghinea@brunel.ac.uk
[*]Corresponding author

ABSTRACT: Creativity is an important ability of an individual to meet the challenges of the 21st century. For this reason, creativity development received priority attention of scholars in the field of education. This special issue collected research articles on innovative theoretical perspectives and original applications related to creative learning in authentic contexts with advanced educational technologies. We received 36 articles and 6 of them were included in this special issue after several rounds of rigorous reviews. In this editorial note, we discuss the background for the special issue and quality management. In addition, we briefly introduce each article selected for the special issue.

Keywords: Creative learning, Authentic contexts, Advanced educational technologies

1. Introduction

Creativity is defined as the ability to produce work that is original and useful (Rhodes, 1987). Produced creative work can be both intangible such as an idea and tangible such as an essay (Sternberg & Lubart, 1999). Scholars suggest that creativity relates not only to the product that results from creative activity but also to the person who creates it, the cognitive processes involved in the creation of the product, and the environmental influences (Mayer, 1989; Rhodes, 1987). Creativity is considered as the most important 21st century skills and is a critical component of any learning program (Bryant, 2010; Lin et al., 2020; Rhodes, 1987; Shadiev et al., 2017a; Sternberg & Lubart, 1999) because creative learning helps learners be innovative, learn new things, try out new ideas, and have new ways of thinking and problem-solving. For this reason, scholars conclude that creativity is important ability in today's world of innovations and creative performance needs to be facilitated in all academic levels (Lin et al., 2020; Shadiev et al., 2022; Shayakhmetova et al., 2020).

Authentic learning environments play crucial role in promoting creative skills development in learners (Davies et al., 2013; Hwang et al., 2019; Jindal-Snape et al., 2013). An authentic environment here is defined as an environment that "preserves the complexity of the real-life context with rich situational affordances" (Herrington & Oliver, 2000, p. 180). Authentic learning environments contains a wide range of available resources that may stimulate learner creativity and make use of such resources supports the growth of ideas (Lin et al., 2020; Shadiev et al., 2022). Furthermore, authentic learning environments give learners greater freedom for imagination, provide rich contexts for the purpose of discovering learner schemas and interests (Wu et al., 2016). Scholars suggested that authentic contexts reflect the way that the knowledge will be used by learners in their real life (Herrington & Oliver, 2000; Shadiev et al., 2017b). Therefore, it is important to encourage creative learning in authentic learning environments.

Creative learning in authentic contexts can be supported by advanced educational technologies (Huang et al., 2017; Hwang et al., 2021; Shadiev et al., 2017a; Shadiev et al., 2015). Advanced educational technology here can be defined as a combination of the processes and tools involved in addressing educational needs and problems, with an emphasis on applying the recent and advanced tools such as computers and other electronic devices (Cifuentes et al., 2011). For example, several advanced educational technologies were listed in Brown et al. (2020), Hwang et al. (2022) and Shadiev and Yang (2020) among them were social networking, artificial intelligence, virtual and augmented reality, robots and many others. Advanced educational technology has many advantages such as it can be used for simulating and restoring some special learning scenes vividly or extending classroom learning to the outdoor environment, enabling learner interaction with the instructor, peers, and learning content (Huang et al., 2017; Shadiev et al., 2017b; Wang, 2020; Wu, 2014). In addition, the technology allows learners to create their own multimedia learning content, share it with their classmates and the instructor, and discuss its strong and weak points thus facilitating creative learning and learner ownership and autonomy (Ahn & Lee, 2015; Huang & Huang, 2015; Shadiev et al., 2017a).

Although many studies have considered the applications of advanced educational technologies to support learning programs, there are not so many studies that focus on creativity. Therefore, there is a need to propose new ideas related to creative learning in authentic contexts with advanced educational technologies, which considers various theories, approaches, techniques, methods, and processes. The aim of this special issue is to collect innovative theoretical work and original applications related to technology-supported creative learning programs in authentic contexts. This special issue focuses on learning models and theories that explain this important dimension, their applications for creative learning in authentic contexts and evidence of their effectiveness based on systematic or empirical data. This special issue also brings research on novel technologies design and on their educational applications that bridges the innovation, pedagogy and practice in technology-supported creative learning.

For this special issue, initially we received 36 submissions from different countries and territories. After that, they were reviewed by well-known international experts in the field. Every article was reviewed by at least three reviewers. After several rounds of a rigorous review process, the best six articles that represent the highest quality suitable for such prestigious journal as *Educational Technology & Society* were selected for inclusion in the special issue. The selected articles address original scientific contributions in the form of theoretical and experimental research and case studies that apply new perspectives on creative learning in authentic contexts with advanced educational technologies.

The first article of the present special issue is *Authentic Learning, Creativity and Collaborative Digital Storytelling: Lessons from a Large-Scale Case-Study* prepared by Nicoletta Di Blas. The author explored whether PoliCultura, a collaborative digital storytelling program for K-12 schools, can foster creativity. All the "stories" submitted to the competition in 2020 were analyzed using a literature-based creativity rubric. The key factors for promoting creativity were discovered and relevant guidelines for educators and researchers were proposed by the author based on the results of the study.

In the second article titled *Open-Ended Tasks Promote Creativity in Minecraft* by Yue Fan, H. Chad Lane and Ömer Delialioğlu, the authors studied the extent to which an open-ended task influences subsequent problem-solving behaviors in a virtual environment. To this end, the authors explored creativity and its relationship with task design in Minecraft and compared a well-defined task group, instructed to follow step-by-step directions, with a group pursuing an open-ended task requiring a higher degree of agency.

The third article by Hyo-Jung Kim, Hyo-Jeong So and Ju-Yeon Park titled *Examining the Effect of Socially Engaged Art Education with Virtual Reality on Creative Problem Solving* focuses on investigating the effect of socially engaged art education with virtual reality on creative problem solving. The participants of the study took part in a four-stage socially engaged art educational program such as appreciation and interpretation of artwork about social issues, discussion on the potential solution to the selected social issue, creating a 3D virtual world to express proposed solutions, and experiencing and sharing 3D virtual worlds. Then participants creative problem-solving skills in three areas such as higher-order thinking, divergent thinking, and problem-solving were examined.

Jin Xinquan, Qiang Jiang, Xingzhu Pan and Wei Zhao in the fourth article titled *The Design and Evaluation of Self-Directed Learning Environment for Creativity Performance* designed an online self-directed learning environment (OSDLE) to improve students' creativity performance. OSDLE helped students plan their learning, learn instructional content, and evaluate and reflect on their learning. The authors carried out the experiment to measure and compare creativity performance of students who learned in OSDLE with that of students who learned in the traditional classroom.

The fifth article is *Does Motivational Design Matter? Motivating Learners in an Augmented Astronomy App* and it was prepared by Chia-Chen Chen, Hong-Ren Chen and Ting-Yu Wang. In the article, the authors proposed creative situated learning via augmented reality (AR), and they developed an AR-based Cosmos Planet Go App to simulate the motion of planets in the universe. With such approach, the authors aimed to help their students better understand the characteristics and features of each planet through its simulated motion in the universe. A quasi-experimental design was applied to demonstrate the effectiveness of the intervention on students learning outcomes.

In the sixth article by Wei-Shan Liu and Ting-Ting Wu titled *Authentic Learning in a Second-year Elementary School Curriculum: Use of a Self-Driving Vehicle for Discussing Innovative Applications of Driverless Cars*, the authors designed authentic learning activities in which they designed authentic learning activities and introduced remote-control cars in order to improve creative thinking skills and problem-solving abilities of elementary

school students. An experiment was carried out to test the effectiveness of proposed intervention on creative thinking skills and problem-solving abilities.

References

Ahn, T. Y., & Lee, S. M. (2015). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778-786. https://doi.org/10.1111/bjet.12354

Brown, M., McCormack, M., Reeves, J., Brook, D. C., Grajek, S., Alexander, B., Bali, M., Bulger, S., Dark, S., Engelbert, N., Gannon, K., Gauthier, A., Gibson, D., Gibson, R., Lundin, B., Veletsianos, G., & Webber, N. (2020). *2020 EDUCAUSE horizon report: Teaching and learning edition*. EDUCAUSE Publications. https://library.educause.edu/-/media/files/library/2020/3/2020_horizon_report_pdf.pdf

Bryant, C. (2010). A 21st-century art room: The remix of creativity and technology. *Art Education*, *63*(2), 43-48. https://doi.org/10.1080/00043125.2010.11519061

Cifuentes, L., Maxwell, G., & Bulu, S. (2011). Technology integration through professional learning community. *Journal of Educational Computing Research*, *44*(1), 59-82. https://doi.org/10.2190/ec.44.1.d

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments in education—A Systematic literature review. *Thinking Skills and Creativity*, *8*, 80-91. https://doi.org/10.1016/j.tsc.2012.07.004

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, *48*(3), 23-48. https://doi.org/10.1007/bf02319856

Huang, Y. M., & Huang, Y. M. (2015). A scaffolding strategy to develop handheld sensor-based vocabulary games for improving students' learning motivation and performance. *Educational Technology Research and Development*, *63*(5), 691-708. https://doi.org/10.1007/s11423-015-9382-9

Huang, Y. M., Shadiev, R., Sun, A., Hwang, W. Y., & Liu, T. Y. (2017). A study of the cognitive diffusion model: Facilitating students' high level cognitive processes with authentic support. *Educational Technology Research & Development*, *65*(3), 505-531. https://doi.org/10.1007/s11423-016-9475-0

Hwang, W. Y., Hariyanti U., Chen, N. S., & Purba, S. W. D. (2021) Developing and validating an authentic contextual learning framework: Promoting healthy learning through learning by applying. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.1876737

Hwang, W. Y., Hoang, A., & Lin, Y.-H. (2022). Smart mechanisms and their influence on geometry learning of elementary school students in authentic contexts. *Journal of Computer Assisted Learning*, *37*(5), 1441– 1454. https://doi.org/10.1111/jcal.12584

Hwang, W. Y., Purba, S. W. D., Liu, Y., Zhang, Y., & Chen, N. S. (2019). An investigation of the effects of measuring authentic contexts on geometry learning achievement. *IEEE Transactions on Learning Technologies*, *12*(3), 291-302. https://doi.org/10.1109/tlt.2018.2853750

Jindal-Snape, D., Davies, D., Collier, C., Howe, A., Digby, R., & Hay, P. (2013). The impact of creative learning environments on learners: A systematic literature review. *Improving Schools*, *16*(1), 21-31. https://doi.org/10.1177/1365480213478461

Lin, L., Shadiev, R., Hwang, W. Y., & Shen, S. (2020). From knowledge and skills to digital works: An application of design thinking in the information technology course. *Thinking Skills and Creativity*, *36*, 100646. https://doi.org/10.1016/j.tsc.2020.100646

Mayer, R. E. (1989). Cognitive views of creativity: Creative teaching for creative learning. *Contemporary Educational Psychology*, *14*(3), 203-211. https://doi.org/10.1016/0361-476x(89)90010-6

Rhodes, M. (1987). An Analysis of creativity. In S. G. Isaksen (Ed.), *Frontiers of creativity research: Beyond the basics* (pp. 216–222). Bearly.

Shadiev, R., Huang, Y. M., Hwang, W. Y., & Liu, T. Y. (2017a). Cognitive diffusion model: Facilitating EFL Learning in an authentic environment. *IEEE Transactions on Learning Technologies*, *10*(2), 168-181. https://doi.org/10.1109/tlt.2016.2574356

Shadiev, R., Hwang, W. Y., & Huang, Y. M. (2017b). Review of research on mobile language learning in authentic environments. *Computer Assisted Language Learning*, *30* (3-4), 284-303. https://doi.org/10.1080/09588221.2017.1308383

Shadiev, R., Hwang, W. Y., Huang, Y. M., & Liu, T. Y. (2015). The impact of supported and annotated mobile learning on achievement and cognitive load. *Educational Technology & Society*, *18*(4), 53–69. https://www.jstor.org/stable/jeductechsoci.18.4.53

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

Shadiev, R., Wang, X., Liu, T. Y., & Yang, M. (2022). Improving students' creativity in familiar versus unfamiliar mobile-assisted language learning environments. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2021.2023891

Shayakhmetova, L., Mukharlyamova, L., Zhussupova, R., & Beisembayeva, Z. (2020). Developing Collaborative Academic Writing Skills in English in CALL Classroom. *International Journal of Higher Education*, *9*(8), 13-18.

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Creativity research handbook* (pp. 3–15). Cambridge University Press. https://doi.org/10.1017/cbo9780511807916.003

Wu, T. T., Huang, Y. M., & Shadiev, R. (2016). The Application of multi-dimensional learning portfolios for exploring the creativity learning behavior in engineering education. *Proceedings of the ASEE Annual Conference and Exposition, ASEE 2016*. https://doi.org/10.18260/p.27355

Wang, Y. H. (2020). Integrating Games, e-Books and AR Techniques to Support Project-based Science Learning. *Educational Technology & Society*, *23*(3), 53–67. https://www.jstor.org/stable/26926426

Authentic Learning, Creativity and Collaborative Digital Storytelling: Lessons from a Large-Scale Case-Study

Nicoletta Di Blas
HOC-LAB, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy // nicoletta.diblas@polimi.it

ABSTRACT: PoliCultura is a collaborative digital storytelling program for schools (K-12), which has gathered so far more than 41,000 students. It is an example of authentic learning experience, for a number of reasons, including the fact that students are required to do a "professional" work that will be made public in the frame of a competition. The paper investigates whether PoliCultura can foster creativity, by analyzing all the "stories" submitted to the competition in 2020 using a literature-based creativity rubric, with positive results. The analysis of the teachers' pedagogical reports sheds light on the key factors for promoting creativity, which are, for the teachers: to be a facilitator of the activity, to promote collaboration, to "open up" to external support and stimuli beyond the classroom and to foster the students' individual talents. Guidelines for designers of educational tools are drawn too: provide a clear path while at the same time allowing a wide degree of freedom, keep the threshold low and plan for the teacher to be at the center-stage.

Keywords: Authentic learning, Creativity, Digital storytelling

1. Introduction

To argument in favor of why creativity should be "taught" at school is not necessary. It has been claimed far and wide, in connection to our "VUCA" world (characterized by Volatility, Uncertainty, Complexity, and Ambiguity) that seems to require, more than ever in history, the capability to adapt, solve problems and find solutions that are "Novel, Effective and Whole": "the emphasis on creativity has never been as pressing, or as academically discussed, as it is in present day" (Mishra et al., 2013, p. 10); in the same line Ohler (2013, p. 13) says that "success in the digital age, both personal and professional, lies in understanding that digital technologies provide one of the greatest imagination creativity amplifiers humankind has ever designed." Similarly, to argument in favor of authentic learning as a trigger for creativity is "not so" necessary, though further evidence may be welcome (Davies et al., 2013; Jindal-Snape et al., 2013). *What we are in need of are ways to foster creativity in our students*: ways to "teach" how to be creative, guidelines to design educational experiences (and tools) that put creativity at center stage. As Henriksen et al. (2015, p. 458) say: "along with this shift toward increased creativity in the classroom comes a need for student project work and class assignments that emphasize creativity." In other words: we know creativity is important, we know authenticity can trigger it: but how can we teach this in our classrooms?

In order to answer this question, the study presented in this paper takes an unusual approach: instead of setting up an experiment, it is based on an *in vivo* study of a whole "living creature": PoliCultura, a program by HOC-LAB at Politecnico di Milano (Italy), that involves groups of students (from K to 12) into collaborative Digital Storytelling (DST). Launched in 2006, PoliCultura is one of the largest deployments of collaborative DST at school in the world: it has involved, so far, more than 43,000 students and 3,200 teachers from 18 countries. The study concludes that PoliCultura is an example of authentic learning activity, fostering creativity, and identifies some key triggering elements that can be taken into considerations by instructors and designers wishing to pursue a similar impact.

After a literature review on creativity, authentic learning and DST in education, the paper describes the case-study and the method of investigation. The results of the interpretation of the program in the light of the authentic learning paradigm, on one side, and creativity, on the other, will lead to the conclusions, in which guidelines for teachers and designers of tech-based educational experiences are drawn.

2. Literature review

The three elements ("Authentic learning," "Creativity" and "Digital Storytelling") this paper intertwines are dealt with in the literature in thousands of books and papers. In what follows, the essentials for each are presented, in relation to this study's focus.

2.1. Creativity and education

Creativity has been identified as a key competence to learning (Kereluik et al., 2013). The standard definition of creativity was developed in the field of psychology (Stein, 1953) as requiring two elements: novelty and utility. This definition is echoed by Robinson (2006), in his TED-talk "Do schools kill creativity?" where he defines creativity as "the process of having original ideas that have value" and underlines the need to "be prepared to be wrong" if you want to "come up with something original." In the same line, Mishra et al. (2013) define creativity as something "Novel" (bringing something into the world that was not there before), "Effective" (useful) and "Whole" (bound to the context within which it occurs): something "NEW."

Creativity in education is seen as something we should not "squander" (again, Robinson, 2006): everyone is potentially creative (Ma et al., 2018) but the school environment in which they are educated influences whether they can develop this potential or not (Giroux & Schmidt, 2004; Nickerson, 2010).

Other scholars focus on what to do (or not) to "teach" creativity. Torrance (1987) identifies 9 ways to teach kids how to think creatively, among which the most relevant for the present case are "creative arts as vehicles" and "motivation, reward, competition," as it will be discussed later. Amabile et al. (1996), as reported in Wadaani (2015, p. 676) "believe that creativity can be encouraged through factors that promote intrinsic motivation with a positive sense of challenge and a focus on the work itself. Such important factors of creativity development […] include autonomy space, creativity encouragement, commitments with clear goals, mutual openness to ideas, and constructive challenge with appropriate reactions and feedback." Wadaani (2015) says that teaching for creativity is not a method but rather a "philosophy" and that teachers "can practice some teaching styles to create an environment of creativity that strengthens the teaching methods that they typically use" (p. 675). Morris (2018) notices how creative learning outcomes are correlated with teachers leaving control to the learners.

A strong relation has been identified between creativity and the infusion of technology into the classroom. As Henriksen et al. (2021, pp. 1, 2) point out, "this rising interest in creativity has occurred during a period of significant societal change due to rapid shifts and developments in technology. Technologies are altering how humans think, work, live, play and create faster than ever. It is, therefore, no surprise that this interest in digital technologies has emerged alongside creativity as critical to contemporary education." The same authors complain that "despite the development of educational policy about creativity, and research into technologies and practices aligned with creativity, instantiating these into the realities of classrooms remains a significant challenge." There is work acknowledging the intersection between these realms (e.g., Sullivan, 2017 or Mehta et al., 2019), but still lacking in the examination of "the creativity-technology relationship from within the classroom, […] *in situ*," while "most of the research in the nexus of creativity, technology and education is conducted from […] outside the experiences of practitioners" (again Henriksen et al., 2021, p. 14, to which the reader is referred for a comprehensive review on this topic).

2.2. Authentic learning

Connecting classroom learning with the outside world is the central point of a great variety of teaching approaches that can be considered declinations of what is more generally defined as "authentic learning." According to Herrington et al. (2014), authentic learning is not a theory of learning, but a guideline to be considered when designing the curriculum. The idea is that students learn more effectively and feel more motivated if they are placed in a context that reflects real life, in contact with relevant issues and achievable projects. In authentic learning, students learn by doing and acquire skills that will be useful in their professional life, like critical thinking, problem solving and research methods. Authentic learning puts students at the center of the stage (Reeves et al., 2002). Students are faced with as-much-as possible realistic problems where the solution is not obvious nor pre-defined and have to take action in order to solve them. Authentic learning highly enhances students' agency due to the strong link with the real world. It tries to amend what Graham (2003) in a quite insightful blog post, complains about: the lack of perception of relevance by students about what is taught them. Authentic learning has sometimes been associated to DST educational activities, especially in the frame of pre-service teachers training (e.g., Chung, 2021; Heo, 2011; Sadik, 2008), where "authenticity" refers mainly to their professional future use in the classroom.

2.3 Digital Storytelling in education

2.3.1. The origins

DST is a practice born in the late 90s in California at a Center first located in San Francisco then moved to Berkeley and named "Story Center" (www.storycenter.org). The Berkley center offers workshops to empower personal storytelling ("life stories") through multimedia (https://www.storycenter.org/history). Since then, DST has experienced a huge success in various fields, like for example cultural heritage or corporate communication, and especially in education.

One of the first examples of educational usage of DST is reported by Mellon (1999), who asked college students to develop a story about a person in their family who had had an impact on their life and acknowledged the motivation the activity raised. This experience was in line with the original DST "philosophy" that puts self-expression at the center of the stage (Burgess, 2006; Meadows, 2003). In 2005 Robin and Pierson, scholars at the University of Houston (home to an important center for DST), matched a digital photography course to a storytelling course, again noticing how the activity was capable of enhancing the students' motivation. In a many-times re-edited book, Lambert and Hesslers (2018) explain the "7 steps" for creating a digital story, thus forging a tool for educators to embed digital storytelling in the classroom. Again Robin (2007) sketched educational experiences in which students would receive assignments that would first require them to carry out research on a topic and then to choose a particular point of view under which to describe it. He underlines how this process can capitalize on the students' creative talent that is expressed when they are asked to do research (learning how to use libraries and the internet to find rich and deep content), to analyze it and synthesize it and finally tell their own version of the story (Robin, 2016). Again, the emphasis is on the personal point of view. This trend, which characterized the origin of DST, reaches to our days: for example, Kim and Li (2021, p. 33), while describing an experience with DST in middle-school, pinpoint "how students expressed their voices, identities, and emotions using the multimodal resources available in digital stories." It must be noted, though, that storytelling has broadened its scope, from sheer self-expression to other themes. Robin himself (2008, p. 224) points out how digital storytelling can obtain excellent educational results when proposed to create, for example, historical documentaries.

2.3.2. Individual vs. collaborative digital storytelling

The emphasis on a personal point of view brings about that, most of the times, stories are done by individual students rather than by groups, even if peer-to-peer sharing is praised as an occasion to introduce a collaborative component in the activity (again, Robin, 2007). Similarly, Gresham (2014, p. 52) implements a creative writing activity where each student is asked to write her/his own story, and notes how collaboration emerges as a relevant aspect: "in time the boys began to see collaboration as something beyond working together and sought inspiration from each other. […] They found themselves sharing ideas on the topic beyond the structure of the classroom, which indicated that they were engaged in their creative writing." Examples of collaborative DST, where students are co-creators of a "story," can be found in the field of Human-Computer-Interaction and outside the scope of formal education, mostly involving young children (preschoolers/first graders) typically playing with "objects" that help making up the story, rather than writing a full script (see Di Blas et al., 2012, for a comprehensive review). As far as formal education goes, collaborative digital storytelling has been explored as a strategy to enhance the learning of a new language, especially English (Nguyen et al., 2020; Nishioka, 2016; Hwang et al., 2016; Chao & Hung, 2014, Hafner & Miller, 2011), or as a playful way to encourage literacy at primary school level (Del-Moral-Pérez et al., 2019). Still, be it for the focus of DST (self-expression) or the nature of collaborative technologies ("most tools, environments and interfaces for co-located collaboration are designed to support the interaction of small groups," Di Blas et al., 2012, p. 271), the diffusion of collaborative DST is limited.

2.3.3. Educational benefits

Digital storytelling has been connected to benefits of various kinds: self-expression, communication skills, media literacy and – quite expectedly – creativity (Di Blas & Ferrari, 2014; Gresham, 2014; Nordmark & Milrad, 2012; Ohler, 2013; Schmoelz, 2018). There is a general consensus on the relation between DST and creativity; the very task at stake hints at this skill: students are asked to produce something original, something they call their own; moreover, as Robin (2016) points out, all the activities "tap into other creative talents such as creating [the students'] own visual images, taking photographs for their stories and adding colors, transitions and recorded narration" (p. 20).

This paper presents an educational experience that lies at the crossroad of creativity, authenticity and DST. In line with the literature, it fosters creativity "through factors that promote intrinsic motivation with a positive sense of challenge and a focus on the work itself" (Amabile et al., 1996, p. 1161). The flavor of authenticity is mainly provided by the fact that the final "product" is submitted to a competition and made public. As regards the approach to DST, PoliCultura is one of the few examples of large-scale programs of collaborative DST, where a whole group of students is asked to create a story together. The main contribution of the study is that, as advocated in the literature, it "gives voice to practitioners" and draws guidelines on how to design authentic, creativity-triggering educational experiences.

3. The case-study

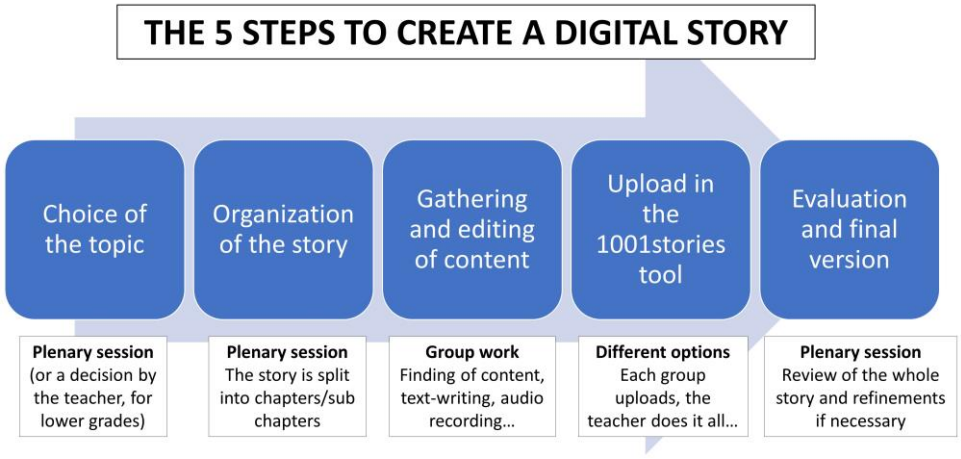
PoliCultura is a collaborative digital storytelling competition designed and run since 2006 by HOC-LAB, a laboratory in the Department of Electronics, Information and Bioengineering at Politecnico di Milano (the largest technical university in Italy). How does the program work? A call is launched at the beginning of the school year, in October. Groups of students, under the guidance of one or more teachers, can take part. All levels, from K to 12, are welcome. Teachers sign up for their teams and are given access to an authoring tool (called "1001stories") with which to create the digital story and to a Massive Online Open Course (MOOC) on digital storytelling in education (for the teachers only). Scardamalia and Bereiter (2006, p. 115) definition of a "knowledge building technology" perfectly fits the 1001stories tool: "a knowledge building technology should facilitate using information, as distinct from learning it. Obtaining, recording, and storing information would become subsidiary functions, designed to serve purposes of knowledge creation." 1001stories is a sophisticated technology in itself, but almost transparent for the user, with a very low threshold of technical knowledge required (Resnick & Robinson, 2017): it is like a pencil, with which anything can be written.

Teams have time from October to March to complete their work and submit it to the competition: the average completion time is a month. It is difficult to estimate how much time is needed to create the digital story, since a lot depends on the choices the team makes concerning the various parts of the work. Based on the feedback by teachers, 20 hours seem to be the minimum. At the end of the work, teachers are asked to submit a report on the pedagogical experience, describing its rationale (see Appendix 4). Once the "story" is submitted, three rounds of jury take place in order to decide who the winners are. At the end, an awards' ceremony takes place at the university's premises, during which the winners are unveiled (Figure 1): the ceremony is attended by almost 300 participants and followed in streaming by thousands of people. All the works are made public in the competition's website (www.policultura.it).

Figures 1. The awards' ceremony, on the left, and on the right a digital story done with 1001stories (by pre-school students in Livermore, CA)

Teachers are provided with instructions on how to create a digital story, which are reinforced during the MOOC: they are quite loose, to allow maximum freedom of expression and for placing the teacher at the center-stage of the process, as orchestrator of the activity. 5 basic steps are suggested (Figure 2).

Figure 2. The 5 steps to create a digital story, with the main organizational strategies.



- Choice of the topic
  Teams can tackle whatever topic they prefer, interpreting the term "story" in a quite broad sense (in line with the literature: see Robin, 2008). Over the years, "stories" about any possible topic, ranging from local cultural heritage to school subjects (such as physics, geography, math…) or school outings have been created. In any case, the teacher makes sure the activity is connected to the curriculum and/or the competences the school is bound to foster.

- Organization of the story
  The topic is organized into chapters and (not mandatorily) sub-chapters. This usually goes with splitting the team into small groups, each in charge of part of the story.

- Content: research and production
  Students look for the content they need, doing "traditional" desktop research but also going beyond the school's boundaries involving experts, families and others (Di Blas & Paolini, 2013a). They edit the material in various ways: they write scripts so that they are appropriate for orality (e.g., making sure to repeat some concepts) and an interactive access (i.e., making sure each piece of the story is reasonably independent), they edit videos and images, record audios, select the proper music, decide on the interface labels, etc. This is a phase in which other technological tools, apart from the 1001stories authoring tool, are used: video and image editing apps, for example. In some cases, teams use a cloud environment for sharing the content and thus facilitating the group work.

- Upload in the authoring tool
  All the materials prepared in phase 3 are uploaded in the 1001stories authoring tool.

- Evaluation and final version
  The output of phase 4 is critically analyzed, in a plenary session: amendments and redesign are put into place to create the final version.

All the above phases undergo different "interpretations" according to the school level (e.g., the decision on what topic to deal with is usually taken by the teacher at level K up to levels 4-5) and also to what benefits the teacher is after (e.g., if she is after cognitive benefits, she will focus on phase 3 and especially on desktop research; if she is after technical benefits, she will equip the students with the content and focus on the technical tasks – phases 3 and 4 – instead). It is highly unlikely for a team to go in depth into all phases and gain all possible benefits in full (Di Blas & Ferrari, 2014). Benefits range from the "traditional" ones (e.g., cognitive benefits related to a better understanding of the topic dealt with) to competences (communication skills, media literacy skills…) and less traditional skills (e.g., professional skills, like "understanding what a deadline is"). Among the benefits there is creativity, which is the focus of this study.

4. Method of the study

As mentioned in the introduction, this study aims at answering the question "how can educational experiences that put creativity at center stage be designed?" and in order to do so it tries to elicit from a "living organism"

like PoliCultura the key elements and triggering factors that make it an authentic and creativity-oriented educational experience. The very first steps therefore are interpreting PoliCultura's main features through the lens of the "authentic learning" paradigm (Reeves et al., 2002) and then investigating whether creativity is actually fostered by the activity; the third step consists in the analysis of the teachers' pedagogical reports, to identify what the recurring elements that characterize the experience are.

Creativity is something that can be perceived, but is quite difficult to assess. In previous years, participants were asked to self-assess their improvement, but the results were positive to the point of raising suspicion: only top scores were given, probably due to the satisfaction of having completed a challenging task. So, it was decided to switch "from the producers to the product," as suggested by Henriksen et al. (2015), and to use an adapted version of their rubric that revolves around the three attributes of "Novel," "Effective" and "Whole" (see Appendix 1).

In June 2020, 58 "stories" by teams of 20-25 students each (1,160-1,450 students on the whole) were submitted to the competition. They were analyzed, independently, by three reviewers with years-long experience with DST in education and who have been part of the referees' board of the competition since 2006. In order to ensure inter-reliability of evaluation, a number of collaborative sessions among them were held, to compare the scoring "style," fine-tune the use of the rubric and set a common standard. Reviewers independently scored the digital stories on a Likert scale from 1 to 5, where 5 was the most positive value. Hereafter, the definitions for the lowest and top scores are reported.

- **Novel**:
  - Lowest score (1): Complete lack of anything unique or novel, lack of content and substance to offer opportunities for novelty (e.g., a standard story, copying existing models).
  - Top score (5): Strong qualities of uniqueness and exciting or interesting to viewers. Is very novel or different from other examples (i.e., other digital stories previously submitted to the competition).
- **Effective**:
  - Lowest score (1): Complete lack of pedagogical effectiveness, and lack of content or substance. A confusing approach, or highly limited presentation of subject that does not make up a coherent picture (e.g., where the contributions by the different groups to the final work are not harmonized).
  - Top score (5): Excellent and highly effective pedagogical approach to communicating the subject. Makes the subject matter clear and comprehensible to most viewers and presents it in interesting and engaging ways that make the subject come alive.
- **Whole**:
  - Lowest score (1): Little or no aesthetic qualities. Poor, or complete lack of, production values, and indicates little or no thought to the design of the learning experience.
  - Top score (5): Excellent or exceptional aesthetic qualities. Flawless or near-perfect production values. Approach provides rich sensory interest (visual, auditory, etc.) for viewers, and all aspects of the design of the digital story are well thought-out to provide an aesthetically cohesive, or "whole" vision that is exciting, thoughtful and stimulating to viewers.

Furthermore, the reports by the teachers were also analyzed in order to identify the most typical aspects that characterize the experiences. The reports follow a schema (see Appendix 4) with some fields to be filled (with no words' limits): their length spans from 1,500 to 2,000 words. They were manually (reporting the data on Excel files) tagged, looking for evidences of: the teacher's role as facilitator, the relevance of group work, the enhancement of the students' specific talents (over having all the students perform all the tasks, regardless of their preferences), the influence of "external resources" and of the spur provided by the visibility of the competition. No score was given to these factors, only their *substantial* (i.e., quite prominent and playing a significant role) presence was reported (1 = present; 0 = absent).

5. Results

5.1. PoliCultura as an authentic learning experience

First of all, let us see if PoliCultura can be labelled an "authentic learning" experience, following the "authentic learning" paradigm (Reeves et al., 2002). Authentic learning experiences are meant to match, as much as possible, real world experiences, in order to foster motivation in the students. First of all, they mimic a professional, all-round approach (different with respect to decontextualized classroom experiences), in which the path that leads to the final result, which is usually a product valuable in its own right, is ill-defined, leaving to the

students the burden of defining task and sub-tasks. This is exactly what happens in PoliCultura, where students are given the task of creating a multimedia communication artifact taking care of all the aspects of its production, as if it were a professional task. Guidelines are offered but they are quite loose, so that they give vent to the most diverse interpretations. This makes the possible outcomes "multiple" in form, style, content… which is another key component of an authentic learning experience. The completion time is quite long (spanning days/weeks and in some cases even months) and requires students to collaborate: again, this is in line with the authentic learning paradigm, which differentiates traditional assignments, which may require a few working hours and are carried on individually, with respect to authentic, professional-like tasks, which are carried on in a team over a long span of time. In the stories produced by the students, the issue at stake is always relevant for the curriculum or the school-life in a broad sense (as the reader can appreciate, by reading the titles and short descriptions in Appendix 3): it requires students to scavenge sources to find the "raw" content (websites, books and articles, experts…) and therefore it pushes them to examine a topic from different perspectives and separate relevant and non-relevant information. The activity is interdisciplinary, being managed most of the times by two or more teachers of different subjects. There is no official assessment: surprisingly enough, most of the teachers don't score their students for their participation in the activity, probably due to the fact that they would not really know how to nor how to isolate specific aspects (e.g., should they grade… creativity? Communication skills? Group work? Enhanced understanding of the subject? … and how?). The assessment, so to speak, comes from the real world, since the competition's referees provide a feedback on the value (and weaknesses) of each story (the reader is referred to Appendix 3 for an example of evaluation by the competition's referees). Eventually, the stories are real communication artifacts as they are all made public in the project's website. In short, PoliCultura seems to correspond well to the various facets that define an authentic learning experience (The reader can find in Appendix 2 the list of all the characteristics of this paradigm put against PoliCultura's main features).

5.2. PoliCultura as a creative experience

As explained in the method section, the 58 "stories" submitted to the competition in year 2020 were analysed independently by three reviewers using an adapted version of the rubric by Henriksen et al. (2015), which revolves around the notions of "Novel," "Effective" and "Whole," and scoring each aspect on a Likert scale from 1 to 5 (see Appendix 3). The average scores and standard deviation are shown in Table 1.

Table 1. The average scores and SD to the "Novel," "Effective" and "Whole" aspects of the 58 stories

	Average score	Standard deviation
Novel	4.53	0.64
Effective	4.28	0.91
Whole	3.91	0.99

It must be noted that the reviewers' assessment of the creativity of the digital stories is quite high and no story was given the lowest score. How can this be possible? First of all, like in the study by Henriksen et al. (2015), provided that participants make "a fair effort, there is usually some kind of moderate degree of quality or effectiveness. This is rather intuitive [since the rubric] defines a 1 as something that is completely lacking." Moreover, compared to previous years, the number of submissions (58) was halved, but of higher quality. The reason is that the works submitted in June 2020 had been done by those that could be labeled as "COVID19-resilient" teams, i.e., teams of students led by proactive teachers who had managed not only to keep on track with their teaching but also to complete the work for the competition.

It is impossible to convey to the reader the full spectrum of topics, styles, strategies put into place by the participants (see Appendix 3 for the list of stories and their short description). There follows an example of top-scoring digital stories in the 3 parameters respectively.

Example 1 – top score on "Novel": Close Encounters of the Third Kind (middle school – Figure 3a)

This is an interdisciplinary work involving geography and science, about the solar system. It is divided into chapters, each devoted to a planet. Groups of 3 students each took care of each chapter, in which – after faking a space mission – they describe the planet from a scientific point of view as if they were reporters from space. The overall plot as well as the specific communication solutions devised by the kids definitely make this work stand out with respect to the standard stories submitted to the program.

Example 2 – top score on "Effective": From persecutions to the Edict of Milan (primary school – Figure 3b)

After 13 years with PoliCultura, the board of referees was not expecting something new in terms of pedagogical use of the authoring tool, but the story "From persecutions to the Edict of Milan" came as a surprise. Why? Because the teacher devised a clever way to use the program to have kids go in depth into the subject. She decided to focus on content gathering (step 3), deliberately giving less emphasis to all the other steps. Most of the students' effort was therefore put on desktop research, based on which they wrote the scripts that they enacted in front of a green screen. Then, in post-production, an image related to the story's topic was added as a background. The result is standard, in terms of communication quality, but ingenious from a pedagogical point of view. On one side, the students are pushed to learn the subject thanks to the intense desktop research, on the other, they feel motivated (by the performance, the product, the competition, …) as well.

Example 3 – top score on "Whole": *"Ubi tu Gaius, ego Gaia"* (high-school – Figure 3c)

A work about "Love and Marriage in Ancient Rome." After a research on contemporary and ancient sources and a thick interaction between the groups and the teacher that went on "24/7, thanks to our common chat," the students created a theatrical performance and turned it into a digital story that hooks the viewer till the end.

Figures 3. (a) Kids faking a space mission to Mars, in their bunk bed and wearing motorcycle helmets (middle-school); (b) kids reading their script "in a catacomb" (primary school); (c) students disguised as ancient romans, enacting a marriage (high-school)

(a) (b) (c)

5.3. PoliCultura's key elements

As explained above, teachers are asked to submit pedagogical reports about the experience, the analysis of which shed light on how the educational experiences were run and what the key aspects were.

The **teacher's role** is, 82% of the times, that of a facilitator. This is quite typical, with a complex activity like PoliCultura that involves different expertise and requires students to be active participants.

Collaboration and **group work** characterize 100% of the experiences: they take different forms, in the sense that sometimes it is kids who organize themselves into groups (especially at higher levels) and sometimes it is the teacher who organizes the groups; sometimes, students are asked to try all the activities, from desktop research and script-writing to the most technical tasks (video-editing, image-editing, …), but most of the times (76%) they are allowed to perform those tasks that best **fit their talents**: for example, those who can draw well are asked to make the drawings, while those who are good at writing work on the scripts. The promotion of individual talents is acknowledged as a motivator (for a discussion of the pedagogical implications of these different approaches, see Di Blas & Paolini, 2013b). In any case, the story is always the result of a collaboration.

The use of **"external resources"** (families, relatives, local experts…), takes place in 93% of cases. One teacher (in a previous round) shared this insight: "in the last few weeks a father came to work with us and this made us feel less lonely and created a warm climate of complicity in the class that did not end within the classroom's walls. From that moment on, we broke all internal hierarchies, bypassed the school bureaucracy, and we transformed the lesson into a true laboratory of ideas and learning. Now that everything is over, we are really proud of what we have done, and we feel almost "empty" without those hours in which teachers turned into pupils, parents became experts, but above all the kids constructed knowledge by themselves." Eventually, the promise of **visibility** and the participation to the competition are considered as powerful motivators by 83% of the teachers.

6. Conclusions

The results of the study confirm that PoliCultura is a collaborative DST experience that can be labelled as "authentic" and is capable of fostering creativity. The program is characterized by teachers that take an active role as facilitator (rather than "sage on stage"), strong collaboration and group work for the students, where talents are enhanced, the opening to external support from families, other colleagues, experts and eventually the visibility provided by the competition as a powerful stimulus.

Some practical implications, for teachers and designers of educational experiences, can be drawn from these results. As regards teachers, lessons are drawn concerning the "enabling factors" that allow designing an authentic experience in which creativity plays a central role.

Lesson 1: be a facilitator. Don't be the one who knows and controls everything, rather, leave room for the students. This is in line with the literature: for example, Morris (2018; 2020) says that optimal results in terms of creativity are obtained by teachers who don't dominate the scene). In a creative activity as PoliCultura, there is no right nor wrong choice (no need to fear to be wrong, in Robinson's words): students can express themselves and discuss the most appropriate strategies with their peers, under the guidance of the teachers.

Lesson 2: have them collaborate. As not only pedagogy (Scardamalia & Bereiter, 2006; Bryant, 2010) but also other fields like for example design thinking (Gero, 1996) highlight, creativity does bloom in collaborative environments where ideas can be exchanged and refined and new solutions emerge, mitigating pitfalls like "design fixation" (i.e., getting stuck on the one solution the designer is more accustomed to) that occur when working on an issue by ourselves (Purcell & Gero, 1996).

Lesson 3: extend your classroom. Go beyond the school's boundaries and involve families, relatives, the local museum, experts… Looking for sources of knowledge and support beyond the school books and the school environment fosters students' agency (Di Blas & Paolini, 2013a).

Lesson 4: identify an external stimulus. Public visibility, even more a competition, are a powerful stimulus to do one's best. As seen in the state-of-the-art section, "competition" (Torrance, 1987) and "positive sense of challenge" can work as creativity-triggering factors.

Lesson 5: promote talents. Contrary to what happens in daily activities, special projects allow putting into place a different strategy (closer to the one we would find in the work place, where everyone is a specialist of something): students are pushed to work on what they are best at rather than doing everything. Leaving aside the (undoubtably fundamental) discussion on whether this is "right or wrong" (the reader is referred to Di Blas, 2013), we must note that having students focus on what they are good at triggers their motivation enormously.

On the educational technology designers' side, these are the lessons' learned.

Suggestion 1: provide loose tracks. Design a tool that at the same time offers a clear path *and* a wide degree of freedom. 1001stories takes you by the hand, so no teacher feels lost, but at the same time it does not place too tight constraints to their freedom of expression.

Suggestion 2: let the teacher do her job. Leave the pedagogy of the experience into the teacher's hands. Let the tool support the "dirty work" and leave to the teacher what she can do best, i.e., the pedagogical organization (group work, individual work, who does what, when, …). Don't design a tool that puts the teacher in a corner.

Suggestion 3: keep it simple. Keep the technology-knowledge threshold low, if you want the tool to be used by schools (Resnick & Robinson, 2017). The technology behind the tool can be quite sophisticated, but it must be easily usable: otherwise, barriers like lack of competences, lack of time to gain them, will hinder its use.

Eventually, some limitations must be acknowledged: first of all, someone may wonder why there wasn't' a control group to fully validate the results of the study. The reason is that PoliCultura is not an experiment, it is a real service offered to thousands of students (between 1,160-1,450 in the 2019-20 round only) in a real context of use with the "myriad particulars" that make the educational science so hard to pursue (Berliner, 2002). Moreover, given its "living organism" complexity, it would be quite difficult to decide how to design the control group: what should the experience be stripped of? Of the stimulus offered by the competition (minimizing the "authenticity" of the activity)? Of the group-work (eliminating collaboration)? Of all technological tasks, having students work on paper? Or of just some of them? The control group experience might look either too similar,

allowing to draw conclusions on just one/few aspects (e.g., stripping the competition, it might be possible to discover something about… motivation?), or so different as to be *de facto* incomparable (e.g., having students work on paper, by themselves, without a competition nor external stimuli…). In other words, in this study the question is not "whether an armadillo is better than, let's say, a fish, but how the armadillo works" (Bolchini et al., 2010). Second, further studies should investigate the different school levels, where surely the teacher's role varies, what students do/don't do varies, the "stories" style and the content vary, etc. The competition's referees have the perception that primary school kids' works are the "wildest" in terms of creativity, but this would need further exploration. Third, which ingredients of the experience are essential or just more important than the others, in order to ensure a "successful" educational experience, is still obscure. Would participants show the same degree of creativity and commitment, without the competition? Would stories be so varied if the tool was more restrictive? Would the educational benefits be so good, if we asked for less (e.g., shorter stories)? These are open questions that are in our research agenda for the years to come.

Acknowledgement

I would like to warmly thank Danah Henriksen and Punya Mishra (Arizona State University) for their invaluable help with the creativity rubric.

References

Amabile, T., Conti, R., Coon, H., Lazenby, J., & Heroon, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal, 39,* 1154-1184. https://doi.org/10.5465/256995

Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational researcher*, *31*(8), 18-20. https://doi.org/10.3102/0013189X031008018

Bolchini, D., Di Blas, N., Paolini, P., & Poggi, C. (2010). Biological life cycle: A New interpretation for the evolution of ICT applications [Paper presentation]. *International Conference "Educational Design Research: Local Change and Global Impact,"* University of Georgia. https://www.academia.edu/3620411/Biological_Life_Cycle_A_New_Interpretation_for_the_Evolution_of_ICT_Applications

Bryant, C. (2010). A 21st-century art room: The Remix of creativity and technology. *Art Education, 63*(2), 43-48. https://doi.org/10.1080/00043125.2010.11519061

Burgess, J. (2006). Hearing ordinary voices: Cultural studies, vernacular creativity and digital storytelling. *Continuum*, *20*(2), 201-214. https://doi.org/10.1080/10304310600641737

Chao, Y. W., & Hung, H. T. (2014). An Exploratory study on university students' perceptions of using digital storytelling to enhance English learning. In T. Bastiaens (Ed.), *Proceedings of World Conference on E-Learning* (pp. 350-355). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/148796/

Chung, M. H. (2021). Digital storytelling project for language teachers: Using personal narratives and technology for authentic learning experience. In M. Dolores Ramírez-Verdugo & B. Otcu-Grillman (Eds.) *Interdisciplinary Approaches Toward Enhancing Teacher Education* (pp. 50-68). IGI Global.

Davies, D., Jindal-Snape, D., Collier, C., Digby, R., Hay, P., & Howe, A. (2013). Creative learning environments in education—A Systematic literature review. *Thinking skills and creativity*, *8*, 80-91. https://doi.org/10.1016/j.tsc.2012.07.004

Del-Moral-Pérez, M. E., Villalustre-Martínez, L., & Neira-Piñeiro, M. D. R. (2019). Teachers' perception about the contribution of collaborative creation of digital storytelling to the communicative and digital competence in primary education schoolchildren. *Computer Assisted Language Learning*, *32*(4), 342-365. https://doi.org/10.1080/09588221.2018.1517094

Di Blas, N., & Ferrari, L. (2014). Digital storytelling at school: What kind of educational benefits? *International Journal of Arts and Technology*, *7*(1), 38-54. https://doi.org/10.1504/IJART.2014.058942

Di Blas, N., & Paolini, P. (2013a). Beyond the school's boundaries: PoliCultura, a large-scale digital storytelling initiative. *Educational Technology & Society*, *16*(1), 15-27.

Di Blas, N., & Paolini, P. (2013b). Technology and group work: Inclusion or diversification of talents? In D. Parmigiani, V. Pennazio, & A. Traverso (Eds.), *Learning & Teaching with Media & Technology. ATEE-SIREM Winter Conference Proceedings* (pp. 218-231). ATEE aisbl.

Di Blas, N., Paolini, P., & Sabiescu, A. G. (2012). Collective digital storytelling at school: A Whole-class interaction. *International Journal of Arts and Technology*, *5*(2-4), 271-292. https://doi.org/10.1145/1810543.1810546

Gero, J. S. (1996). Creativity, emergence and evolution in design. *Knowledge-Based Systems*, *9*(7), 435-448. https://doi.org/10.1016/S0950-7051(96)01054-4

Giroux, H. A., & Schmidt, M. (2004). Closing the achievement gap: A Metaphor for children left behind. *Journal of Educational Change*, *5*, 213–228. https://doi.org/10.1023/B:JEDU.0000041041.71525.67

Graham, P. (2003). *Why nerds are unpopular*. http://www.paulgraham.com/nerds.html

Gresham, P. (2014). Fostering creativity through digital storytelling: "It's a paradise inside a cage." *Metaphor*, (1), 47-56. https://www.englishteacher.com.au/resources/command/download_file/id/162/filename/141Digitalstorytelling.pdf

Hafner, C. A., & Miller, L. (2011). Fostering learner autonomy in English for science: A Collaborative digital video project in a technological learning environment. *Language Learning & Technology*, *15*(3), 68-86. https://doi.org/10125/44263

Henriksen, D., Creely, E., Henderson, M., & Mishra, P. (2021). Creativity and technology in teaching and learning: A Literature review of the uneasy space of implementation. *Educational Technology Research and Development*, *69*(4), 2091-2108. https://link.springer.com/article/10.1007/s11423-020-09912-z

Henriksen, D., Mishra, P., & Mehta, R. (2015). Novel, effective, whole: Toward a NEW framework for evaluations of creative products. *Journal of Technology and Teacher Education*, *23*(3), 455-478. https://www.learntechlib.org/primary/p/151574/.

Heo, M. (2011). Improving technology competency and disposition of beginning pre-service teachers with digital storytelling. *Journal of Educational Multimedia and Hypermedia*, *20*(1), 61-81. https://www.learntechlib.org/primary/p/35353/.

Herrington, J., Reeves, T. C., & Oliver, R. (2014). Authentic learning environments. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 401-412). Springer. https://doi.org/ 10.1007/978-1-4614-3185-5_32

Hwang, W. Y., Shadiev, R., Hsu, J. L., Huang, Y. M., Hsu, G. L., & Lin, Y. C. (2016). Effects of storytelling to facilitate EFL speaking using Web-based multimedia system. *Computer Assisted Language Learning*, *29*(2), 215-241. https://doi.org/10.1080/09588221.2014.927367

Jindal-Snape, D., Davies, D., Collier, C., Howe, A., Digby, R., & Hay, P. (2013). The Impact of creative learning environments on learners: A systematic literature review. *Improving schools*, *16*(1), 21-31. https://doi.org/10.1177/1365480213478461

Kereluik, K., Mishra, P., Fahnoe, C., & Terry, L. (2013). What knowledge is of most worth: Teacher knowledge for 21st century learning. *Journal of Digital Learning in Teacher Education*, *29*(4), 127-140. https://doi.org/10.1080/21532974.2013.10784716

Kim, D., & Li, M. (2021). Digital storytelling: Facilitating learning and identity development. *Journal of Computers in Education*, *8*(1), 33-61. https://doi.org/ 10.1007/s40692-020-00170-9

Lambert, J., & Hessler, B. (2018). *Digital storytelling: Capturing lives, creating community*. Routledge. https://doi.org/10.4324/9781351266369

Ma, X., Yang, Y., Wang, X., & Zang, Y. (2018). An Integrative review: Developing and measuring creativity in nursing. *Nurse Education Today, 62*, 1–8. https://doi.org/ 10.1016/j.nedt.2017.12.011

Meadows, D. (2003). Digital storytelling: Research-based practice in new media. *Visual Communication*, *2*(2), 189-193. https://doi.org/10.1177/1470357203002002004

Mellon, C. A. (1999). Digital storytelling: Effective learning through the internet. *Education Technology*, *39*(2), 46-50.

Mehta, R., Henriksen, D., & Rosenberg, J. M. (2019). It's not about the tools. *Educational Leadership*, *76*(5), 64–69. https://www.ascd.org/el/articles/its-not-about-the-tools

Mishra, P., Henriksen, D., & Group, D. P. R. (2013). A NEW approach to defining and measuring creativity: Rethinking technology & creativity in the 21st century. *TechTrends*, *57*, 10–13. https://doi.org/ 10.1007/s11528-013-0685-6

Morris, T. H. (2018). Vocational education of young adults in England: A Systemic analysis of teaching-learning transactions that facilitate self-directed learning. *Journal of Vocational Education & Training*, *70*, 619–643. https://doi.org/10.1080/13636820.2018.1463280

Morris, T. H. (2020). Creativity through self-directed learning: Three distinct dimensions of teacher support. *International Journal of Lifelong Education*, *39*(2), 168-178. https://doi.org/10.1080/02601370.2020.1727577

Nguyen, T. H., Hwang, W. Y., Pham, X. L., & Pham, T. (2020). Self-experienced storytelling in an authentic context to facilitate EFL writing. *Computer Assisted Language Learning*, 1-30. https://doi.org/10.1080/09588221.2020.1744665

Nickerson, R. (2010). How to Discourage Creative Thinking in the Classroom. In R. Beghetto, & J. Kaufman (Eds.), *Nurturing Creativity in the Classroom* (pp. 1-5). Cambridge University Press. https://doi.org/10.1017/CBO9780511781629.002

Nishioka, H. (2016). Analysing language development in a collaborative digital storytelling project: Sociocultural perspectives. *System*, *62*, 39-52. https://doi.org/10.1016/j.system.2016.07.001

Nordmark, S., & Milrad, M. (2012). Mobile digital storytelling for promoting creative collaborative learning. In *2012 IEEE Seventh International Conference on Wireless, Mobile and Ubiquitous Technology in Education* (pp. 9-16). IEEE. https://doi.org/ 10.1109/WMUTE.2012.10

Ohler, J. B. (2013). *Digital storytelling in the classroom: New media pathways to literacy, learning, and creativity*. Corwin Press.

Purcell, A. T., & Gero, J. S. (1996). Design and other types of fixation. *Design studies*, *17*(4), 363-383. https://cs.gmu.edu/~jgero/publications/1996/pur-ger-design-dstudies.pdf

Reeves, T. C., Herrington, J., & Oliver, R. (2002). *Authentic activity as a model for web-based learning* [Paper presentation]. 2002 Annual Meeting of the American Educational Research Association, New Orleans, LA, USA. https://researchrepository.murdoch.edu.au/id/eprint/7626/

Resnick, M., & Robinson, K. (2017). *Lifelong kindergarten: Cultivating creativity through projects, passion, peers, and play*. MIT press.

Robin, B. (2007). The Convergence of digital storytelling and popular culture in graduate education. In R. Carlsen, K. McFerrin, J. Price, R. Weber, & D. Willis (Eds.), *Proceedings of SITE 2007-Society for Information Technology & Teacher Education International Conference* (pp. 643-650). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/24616/

Robin B. (2008). Digital storytelling: A Powerful technology tool for the 21st century classroom. *Theory into Practice*, *47*, 220-228. https://doi.org/10.1080/0040584080215391

Robin, B. (2016). The Power of digital storytelling to support teaching and learning. *Digital Education Review*, (30), 17-29. https://doi.org/10.1344/der.2016.30.17-29

Robin, B., & Pierson, M. (2005). A Multilevel approach to using digital storytelling in the classroom. In C. Crawford, R. Carlsen, I. Gibson, K. McFerrin, J. Price, R. Weber, & D. Willis (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2005* (pp. 708-716). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/19091/.

Robinson, K. (2006). *Do schools kill creativity? TED Talks.* http://www.ted.com/talks/ken_robinson_says_schools_kill_creativity?language=en

Sadik, A. (2008). Digital storytelling: A Meaningful technology-integrated approach for engaged student learning. *Educational technology research and development*, *56*(4), 487-506. https://doi.org/10.1007/s11423-008-9091-8

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97-118). Cambridge University Press. https://doi.org/10.1007/s11423-008-9091-8

Schmoelz, A. (2018). Enabling co-creativity through digital storytelling in education. *Thinking Skills and Creativity*, *28*, 1-13. https://doi.org/10.1016/j.tsc.2018.02.002

Stein, M. I. (1953). Creativity and culture. *Journal of Psychology*, *36*, 311–322. https://doi.org/10.1080/00223980.1953.9712897

Sullivan, F. R. (2017). *Creativity, technology, and learning: Theory for classroom practice*. Taylor & Francis. https://doi.org/10.4324/9781315765143

Torrance, E. P. (1987). Teaching for creativity. In *Frontiers of creativity research: Beyond the basics* (pp. 189-215). Bearly Limited.

Wadaani, M. R. (2015). Teaching for creativity as human development toward self-actualization: The Essence of authentic learning and optimal growth for all students. *Creative Education*, *6*(07), 669. https://doi.org/10.4236/ce.2015.67067

Appendix 1

Coding Rubric– Novel, Effective, Whole

There follows the adaptation for the PoliCultura competition of the coding rubric on creativity by Henriksen et al. (2015). The main reason for adapting the rubric was that it had been developed to evaluate products by prospect teachers, who were asked to design mainly lesson plans and educational projects, while in the case of PoliCultura the products are digital stories done by the students. Therefore, sentences like "…relatively standard approach *to the teaching of the subject matter*" were modified into something like "…relatively standard approach *to the creation of a digital story*," to account for the specificity of the study. It must be noted though that the substance of the rubric was not affected. In what follows, the parts in italic highlight the differences with respect to the original version (which are, as the reader will be able to appreciate, minimal).

Novel

1 – Complete lack of anything unique or novel, lack of content and substance to offer opportunities for novelty *(e.g., a standard story, copying existing models).*

2 – Fairly lacking in unique, fresh or novel qualities. Most elements are quite standard and unconventional.

3 – Relatively standard approach to the *creation of a digital story*. While there may be a few unique qualities, it does not necessarily stand out among other *digital stories.* Average.

4 – Some qualities of uniqueness, and relatively interesting to *viewers.* While aspects may bear certain similarities to standard approaches to *digital storytelling*, it also contains some interesting, fresh or novel qualities.

5 – Strong qualities of uniqueness and exciting or interesting to viewers. Is very novel or different from other examples *(i.e., other digital stories previously submitted to the competition).*

Effective

1 – Complete lack of *pedagogical* effectiveness, and lack of content or substance […]. A confusing approach, or highly limited presentation of subject that *does not make up a coherent picture (e.g., where the contributions by the different groups to the final work are not harmonized).*

2 – Fairly ineffective approach to or presentation of subject matter […]. May have elements that are somewhat boring, confusing, dry, light on content, or do not sufficiently communicate the subject matter clearly to *viewers.*

3 – Somewhat effective *pedagogical* approach to *digital storytelling*, in that some elements of the approach to or presentation of content work well to communicate the ideas clearly in interesting ways. However, there remain some flaws or areas that lack, or that appear to *communicate* the content less successfully. Average

4 – Effective *pedagogical* approach to *digital storytelling*. Clear, thoughtful and interesting approach to *communicating* the content successfully. Little room for misconceptions or confusion – a coherent approach that appears to lead to solid *and coherent picture*.

5 – Excellent and highly effective *pedagogical* approach to *communicating* the subject. Makes the subject matter clear and comprehensible to most *viewers* and presents it in interesting and engaging ways that make the subject come alive.

Whole

1 – Little or no aesthetic qualities. Poor, or complete lack of, production values, and indicates little or no thought to the design of the learning experience.

2 – Few aesthetic qualities, showing weakness in appeal or production values. Clear flaws or minimal thought given to the design *of the digital story*.

3 – Some aesthetic qualities, but also somewhat conventional or standard aesthetic appeal. Some thought to the design *of the digital story* is evident, though overall the production values and aesthetic appeal are fairly average. Reasonably well done, but lacking in any "stand out" appeal.

4 – Good aesthetic qualities, and sharp or polished production values. Approach provides some sensory interest (visual, auditory, etc.) for students, with clear thought to the design *of the digital story*. The aesthetics qualities help make *the digital story* interesting and thought-provoking to *viewers*.

5 – Excellent or exceptional aesthetic qualities. Flawless or near-perfect production values. Approach provides rich sensory interest (visual, auditory, etc.) for *viewers*, and all aspects of the design *of the digital story* are well thought-out to provide an aesthetically cohesive, or "whole" *vision* that is exciting, thoughtful and stimulating to *viewers*.

Appendix 2

PoliCultura as an authentic experience

Table 2. PoliCultura and the characteristics of an "authentic" educational experience

Authentic learning element	How it is described in the literature	How it applies to PoliCultura
Real-world relevance	Activities match as nearly as possible the real-world tasks of professionals in practice rather than decontextualized or classroom-based tasks.	PoliCultura asks students to create a "professional" multimedia product, which is something quite different with respect to what "digital natives" normally do. It means being able to organize the structure of an interactive product, to write scripts for orality and interactivity (using a different syntax, avoiding cross-textual references…), to combine verbal and visual communication…
Ill-defined	Activities require students to define the tasks / sub-tasks needed to complete the activity.	Participants are provided with loose-end instructions, allowing plenty of space for invention. Everything is free, from the topic to deal with, to the structure of the story, the kind of content (slideshow or videos, music or audio…).
Complex, sustained tasks	Activities are completed in days, weeks, and months rather than minutes or hours. They require significant investment of time and intellectual resources.	PoliCultura provides participants with a time frame spanning 5 months to submit their work; the average time needed to complete the story is one month. The task involves different kinds of expertise and resources.
Multiple perspectives	Provides the opportunity for students to examine the task from different perspectives using a variety of resources, and separate relevant from irrelevant information.	This aspect is typical of PoliCultura, where a "system of resources" comes into play for completing the work. The work on the content selection requires to be able to separate meaningful from meaningless sources.
Collaborative	Collaboration is integral and required for task completion	PoliCultura is one of the rare examples of collaborative digital storytelling at school. Everything is done collaboratively and the work is organized around groups of students in charge of specific parts of the story or tasks; they need to orchestrate their effort to create a common "story."
Value laden	Provide the opportunity to reflect and involve students' beliefs and values.	The extent to which students are allowed to put forth their points of view, beliefs and values is up to the teacher(s), but this is the case, most of the times, since telling a digital story is typically related to self-expression.

Interdisciplinary	Activities encourage interdisciplinary perspectives and enable learners to play diverse roles and build expertise that is applicable beyond a single well-defined field or domain.	This is certainly true for PoliCultura. In most of the cases, the work is guided by two or more teachers of different disciplines; learners play different roles (e.g., the group leader, the content editor, …) and these roles rotate among the participants.
Authentically assessed	Assessment is seamlessly integrated with learning in a manner that reflects how quality is judged in the real world.	Most of the time students are not scored on their participation (even if their commitment is quite high), maybe due to the fact that it is quite difficult to assess something so complex and unusual. The assessment, in a sense, is provided by the "real world," i.e., the competition's referees that evaluate the product.
Authentic products	Authentic activities create polished products valuable in their own right rather than as preparation for something else.	All the stories submitted to the competition are whole and complete; they are made public in the project's website.
Multiple possible outcomes	Activities allow a range and diversity of outcomes open to multiple solutions of an original nature, rather than a single correct response obtained by the application of predefined rules and procedures.	There is no "right" story nor "right" way of creating it. Participants are not asked to copy a model but to surprise the referees with ever different interpretations of what the tool can make. The almost 2,000 stories submitted this far are different in terms of style, topic, communication strategy… to the point that they are incomparable.

Appendix 3

The digital stories and their scores

In the following table, the 58 digital stories that were analyzed are presented, together with their scores (on a Likert scale, from 1 to 5, where 5 is the most positive score) by the 3 experts, according to the creativity rubric (Novel, Effective, Whole). SG stands for "School Grade," K stands for "Kindergarten," P stands for "Primary school," M stands for "Middle school" and "H" stands for "Highschool." In addition, a short excerpt from the teacher's report is presented: the reading of a number of them can provide the reader with the "flavor" of the experience.

Table 3. The 58 stories: description, excerpt from the teachers' reports and scores by the reviewers

#	Title and short description of the story	Short excerpt from the teacher's report	SG	N	E	W
1	*Agenda 2030: the world is in our hands* The keyword "respect" is the fil rouge than makes these very young students understand what a healthy relation with the environment and society can be.	"The second step was to open up to the world. Mrs. S.R., mother of one of our children, came to our aid to tell us about her experience as an orphan in a distant land. The fact that not all children have the same standard of living has been the inspiration to bring a distant idea close."	K	5 5 4	4 5 4	4 4 3
2	*Emotions in the magical world of Oz* Through the story of the "Wizard of Oz" by L. Frank Baum important issues like self-esteem, self-confidence, diversity and solidarity are addressed.	"For communicating and playing with the deaf companion, the children have learnt the sign language and with extraordinary speed the necessary skills to welcome and include her."	K	4 3 3	3 2 3	3 2 3
3	*Sophie's magic* A story about a little girl and an inclusive school, where obstacles that sometimes prevent full participation in the social and	"Seeing Sophie with the other children is wonderful, both for the happiness she feels when she is with them and for the empathy, tenderness and care that her companions show to her."	K	5 5 4	5 5 4	2 3 3

	educational life are removed.					
4	*Scientists to the rescue!* A "story" about scientific thinking and more specifically the concept of time, the transformation of matter (ice-water-steam), the concept of liquid.	"We worked on the development of scientific thinking. The desire was to create a didactic tool containing experiences with a constructivist approach."	K	5 5 5	2 2 3	4 3 3
5	*Between dream and reality* The children imagine they are going to visit the fearsome ogre in their village's castle. Through storytelling, they are stimulated to develop creativity, imagination and curiosity.	"Our theme for this year is: 'how to pass on'. […] we are committed to designing and activating experiences that help children develop an individual sense of what is precious. We want children to research and acquire special values, which we define as gifts."	K	4 5 5	5 3 5	2 3 4
6	*From yesterday to today, a dip into the monastic experience* A "TV show" about the birth and evolution of Monasticism from the third century AD.	"There was a great spirit of collaboration between the pupils and the teacher. The atmosphere has always been serene, joyful and for everyone it has been an exciting job."	P	5 4 5	3 2 3	3 2 2
7	*From nature to art* A work on visual art connected to the children's locality, with a focus on nature, facilitated by meetings with experts.	"By focusing on some paintings relating to nature and the local landscape by artists, the pupils entered the world of colors, fruits, vegetables and cultural sites in the area. With the collaboration of experts, themed conferences and workshops were held."	P	4 5 5	4 5 5	2 3 2
8	*From persecutions to the edict of Milan* A project that tells the story of Christianity from its origins on the day of Pentecost to the edict of 313 AD issued by the emperor Constantine.	"The purpose of the work is to push students to get to know a curricular topic becoming protagonists of the knowledge acquired and passing it on to other classmates […] in fact, this topic of study was addressed in parallel classes through the vision of the video made by these students."	P	5 5 5	5 5 5	2 3 2
9	*The beginnings of Christianity* This project tells the story of the origin of the Church and its development especially in the first century AD.	"The children, starting from the texts given by the teacher, were divided into working groups, they reworked the contents to create narratives to be rendered in the form of a documentary."	P	5 5 4	5 5 4	2 2 3
10	*Back in time with Saint Paul* The goal of the work is to tell the story of St. Paul trying to frame him as a historical figure, discovering who he was and what he has become for the Church in the very early years of Christianity.	"…creation of a storytelling made as if it were a small documentary. Both the pupils and the teachers actively participated, in a general climate of active cooperation, interchange and natural joy."	P	4 5 5	4 5 4	4 4 5
11	*I stand by Vanessa* A very original "silent book" that tackles the issue of bullying through the story of little Vanessa, revealing the importance of small daily gestures capable of counteracting hypocrisy and indifference.	"Experience has shown that bullying mainly happens at school. Therefore, it is the school's duty to monitor the students, offering an educational path that allows them, from an early age, to recognize their feelings and communicate them, to reflect on the feelings of others, suggesting strategies for coping with the strongest emotions."	P	3 4 5	3 3 5	4 5 4
12	*Fairy tales are dreams that can*	"If fairy tales are dreams, can they	P	5	5	3

	come true The story reports on a multifaceted educational experience where the reading of fairytales is intermingled with meetings with "real" people who have made their dreams come true.	become true? Sometimes, yes! And that's why Niky came to visit us. He told us his story and got interviewed by the children, who turned into journalists. The doctors diagnosed little Niky with chronic asthma and advised his parents to take him to live by the sea. His parents […] decided to build a boat in their backyard. After months of hard work, the Walkyrie schooner of more than 20 meters, thanks to the special transport and navigation of the river Po', reached the sea, where Niky's dream came true."		5 4	2 2	3 4
13	*To live, see, discover Martina Franca* The motto of this work is: "you can read a town like a book: every stone is a page of history." The goal of the story is to introduce little-known places and traditions of the children's hometown to everyone.	"The use of new technologies and teamwork immediately motivated the pupils to engage in the project. The multidisciplinary and interdisciplinary work was carried out during school hours of history, technology, art, innovation and English."	P	3 4 5	5 5 5	3 4 3
14	*More nature, more life!* The story is set in the school garden used not only for recreation, but also as a laboratory full of life and natural elements that children have discovered through exploration, observation, manipulation and gardening.	"In all phases the work was organized trying to leave a lot of space for the children, starting from their natural curiosity and allowing everyone to interact freely. In the various activities, the pupils were divided into heterogeneous groups, chosen by the teacher, where everyone contributed to the whole work."	P	5 4 5	4 2 5	5 4 5
15	*Knowledge and taste* The theme of the story is food. Starting from personal experiences by the students, it aims at promoting a more aware and responsible attitude towards nutrition and food in general.	"The class was divided into heterogeneous groups and the roles were distributed taking into account the preferences of each pupil and personal aptitudes. In some cases an attempt was made to stimulate pupils to try their hand at something new, overcoming fears and lack of self-esteem."	P	5 4 4	5 2 4	5 4 4
16	*We are good together when ...* A story on rights and the value of rules, to learn to live together in an active and peaceful way and to become aware citizens.	"…an attempt was made to act on the child's daily experience, preparing him/her for the convinced and participatory recognition of the rules of democratic life."	P	5 5 4	5 5 4	5 4 4
17	*Humanity on the go* A poetic story about migrants and travels, including the personal experiences of some of the kids' parents and relatives.	"In class we talked about the expectations of the trip, how much you want to leave or not and the reasons behind a trip; […] the children interviewed their parents who had moved to Italy for different reasons."	P	5 5 5	5 5 5	2 2 3
18	*The dream planet* A path of emotional and cultural exploration and awareness raising on sustainability issues, human well-being, environment protection and care of our common home, in the light of the	"In order to develop critical skills in the students, in addition to a self-assessment by each group of their work, the whole class evaluated the whole work: examining if the texts were suitable for oral use, if the length was adequate, if the images	P	4 4 5	4 4 4	4 3 5

	2030 Agenda.	were in adequate number and consistent with the text."				
19	*United we say NO to bullying* A story about bullying and cyberbullying, including the rap song "The nursery rhyme of the repentant cyberbully".	"Role-playing activities were used to represent the bully, the victim and the spectators, with simulations, improvisations, games. […] Each phase of the project was experienced with great enthusiasm by all the pupils. There has never been a moment of fatigue, even when it happened to repeat shots, scenes or other to improve the performance."	P	4 3 5	5 2 4	4 3 5
20	*Walking among the Forts of Messina* A story of discovery and knowledge of the territory. The pupils have learned that Messina (a town in Southern Italy) hides unknown beauties of great historical value.	"Due to the pandemic, it was decided that each student would focus on a specific fort and create a multimedia presentation to share with the other classmates […] In this way ALL the pupils have taken part to the project."	M	5 5 4	5 5 4	5 5 4
21	*At the twenty-second mile of the Via Salaria ... Saint'Antimo…* An ancient narration that begins with the dream of a prince named Anthemio, from which the town where the kids live in is said to have originated…	"From the texts, the testimonies and some interviews with local historians our work has begun. The students worked in separate groups and produced a video. This video reports an interview to the parish priest of the church of Saint'Antimo…"	M	5 5 4	5 4 4	4 5 4
22	*Etna Sustainable Annular Cycle Path* The narration unfolds starting from the proposal to create an annular cycle path around the volcano Etna (southern Italy) with the indication of the route, and the description of the landscapes, places, fauna, flora and also information of a gastronomic and historical nature.	"The base for the work came from different sources: internet, cartographic material, historical and scientific texts. Students worked three school hours a week, three afternoon hours a week, from January to March 6; from April to May the work was carried out remotely, using Google Meet, to a total of about 80 hours."	M	5 5 4	5 5 4	5 5 4
23	*Ecology of a virus* The Coronavirus (personified as a character) tells, in a quite engaging way, all its characteristics, highlighting the lifestyle people were bound to conduct during the lockdown period.	"The work was divided into three phases. (1) information: the students looked for info on the web; (2) operational: production of family interviews as evidence of the current situation; (3) creative: the kids created audio and video."	M	5 4 5	5 4 3	5 5 5
24	*Climate changes* A raising-awareness story about rising temperatures, melting ice, rising sea levels, increasing extreme weather phenomena, desertification, climate-related migration, economic damage, loss of biodiversity.	"The teachers involved in the project believe that making students 'create' is an effective and educational experience. [...] Students, divided into groups, combine images and words to capture the attention both with images and with verbal communication."	M	5 4 5	4 5 4	2 4 3
25	*The great cemetery of Messina* A short story about a graveyard in Messina (Sicily, southern Italy), artistic testimony of the grandeur with which the cult of the dead was observed in the past.	"The kids together carried out a research work on the great cemetery of Messina, a poorly known cultural asset. They listened to the readings chosen by the teacher, during the remote lectures, and searched for the photos to be included in the work on the internet. The kids really teamed	M	4 5 4	4 5 4	4 3 3

		up."				
26	*Close encounters of the third kind* A funny story about fake interplanetary expeditions from which the main characteristics of the planets that make up the solar system emerge.	"The 'transversal' goals are: to be able to find effective solutions to problems encountered in the implementation of the project, starting with the identification of the best technological solutions; to work as a team to achieve a result greater than the sum of the individual contributions."	M	5 5 5	4 3 5	5 5 5
27	*The Moon: shining object of desire* Fantastic tales and expressive reading complemented by drawings made by the children intertwine the red thread (the Moon) that runs through the narrative allowing the free expression of desires and emotions.	"The pandemic, which started at the end of February, changed the initial project, making it also an opportunity to feed the imagination, at a time when reality was not easy to accept, and to involve pupils, from a distance, helping them stay connected thanks to digital tools."	M	4 4 4	5 5 4	4 3 4
28	*Our impossible interviews* A gallery of "portraits" of great women who have distinguished themselves in STEM.	"The students were divided into groups of three/four, according to their choice, considering that part of the work would take place in extra-school hours. Roles have been established: character, interviewer and movie director."	M	4 5 5	4 5 5	5 4 4
29	*Leonardo: an artist always on the go* An imaginary interview to discover a brilliant character through the places he traveled through during his life, an opportunity to find out who he was and what legacy he left us.	"Considering that the aspect to be privileged is orality, the communication genre of the interview was chosen. Furthermore, in this way, all the children had a place in the final product and learned to manage the emotion of expressively reading aloud and, above all, the emotion of recording and listening to their own voice."	M	5 4 5	5 4 5	4 3 5
30	*Manduria and us, today and yesterday* Narrative of love for the students' hometown, with a focus on the natural and cultural beauties of the area and an historical excursus between past and present.	"Goals: (1) encourage a learning process based on the "mixing" of formal and informal sources and on the creative mash-up of pre-existing didactic contents; (2) strengthen identity, self-esteem, comparison with others, recognition, personal growth and acceptance; (3) regulate the level of confrontation and conflict with others by encouraging collaboration, also between pupils and teachers."	M	5 5 4	5 5 4	5 5 3
31	*Margherita among the stars* This story answers the question: can a brilliant and extroverted character like Margherita Hack serve as an exemplary figure and guide life choices?	"The autobiography 'A life among the stars' sparked the idea of synthesizing the life of Margherita Hack through short animated sequences. Building a cooperative product by manipulating content is an interesting goal that can be easily achieved using animation apps."	M	4 5 5	4 5 5	4 4 5
32	*STEM woman worker* Original narration dedicated to an important character of the school: the school's principal. The students interview her about her	"According to the principles of the challenging 'enactive teaching', more than the learning of predetermined contents, the experience of building knowledge – in our case, of a	M	5 5 5	5 5 5	5 5 5

	life and her vocation to STEM and in particular to biochemistry.	citizenship respectful of the female gender – matters."				
33	*Story of an (almost) successful twinning* Territory, history, art, culture are the chapters of a documentary about two "twin" cities in Italy.	"Surely, the work for PoliCultura has made the bond between the children and the teachers involved even stronger."	M	5 4 5	5 4 5	4 4 5
34	*Tale of our "flipped" classroom* Tale of how the students faced the study of science and mathematics through the methodology of the flipped classroom.	"Roles were defined by the teacher, but the attribution took place independently within each group. - web researcher - graphic designer, - text editor, - video editor. Each pupil was in charge of one aspect, but before the delivery the work was reviewed by all members of the group."	M	5 3 5	5 3 5	5 3 5
35	*A planet on its knees* The topic addressed is that of climate change, through the story of Grandma Flo (personification of mother Earth) and her granddaughter, conceived and created by the children with their beautiful drawings.	"We relied on the action-research approach. An attempt was made to promote as much as possible active learning by the pupils, placing them at the center of the teaching activity, while the teacher played the role of tutor in the process of knowledge discovery-acquisition."	M	5 5 4	5 5 4	5 5 4
36	*A year-long earthquake: stories of resistance and resilience under the Mongibello* It was chosen to analyze from a scientific and emotional point of view the event of the earthquake that in December 2018 hit the municipality where the school attended by the children is located.	"Study of earthquakes: it was chosen to address the topic in a different way, not only from a scientific point of view, but also from the point of view of the effects on the lives of people who are affected by an earthquake."	M	5 3 5	5 3 5	5 3 5
37	*The dream school* The narration tells the dream of a new school, conceived and designed starting from the knowledge of how the school attended by the children has changed in the last 60 years.	"We believe that the project has benefited the students who were very willing to deal with the proposed activities. In particular, the assignment of specific roles has highlighted special talents that would have hardly surfaced with a more traditional teaching method."	M	5 5 5	5 5 4	2 3 3
38	*"Ubi tu Gaius ego Gaia"* A work about "Love and Marriage in Ancient Rome." After a desktop research, the students created a theatrical performance and turned it into a digital story that hooks the viewer till the end.	"The focus was on the peer group, which formed a sort of social laboratory, in which to develop dynamics, experiment with activities, plan, share, improving self-esteem and relational and communication skills."	H	5 5 5	5 5 5	5 5 5
39	*3,14159265 and many other stories* The protagonist of the project is the number most invoked by students from all over the world: pi. In the work, properties and stories about it are told.	"Three disciplines were involved in the activity: literature, mathematics and English. The literature teacher followed the part relating to the communication and storytelling techniques. The math teacher developed the topics during the workshops. The English teacher	H	4 5 5	4 5 5	4 5 5

		followed a group that decided to make their part in English."				
40	*A Day in Prison* The story features the dark and tormented atmosphere of a women's prison. It was born out of the desire to fight to live in a better world where the strongest do not prevail and where everyone can exercise their freedom of thought and action.	"Through this project it was possible to rediscover how important it is to learn to know each other, especially among peers and to understand with great surprise the precious value of friendship. Each student managed to share emotions and thoughts, they treasured this experience, transforming it into an opportunity to improve themselves."	H	5 5 5	4 5 5	4 5 3
41	*Water fun!* The story examines the theme of exploitation of water resources and sustainability from multiple points of view, starting with the study of rainfall in the students' home town.	"During the activities, the students were able to experiment with new tools and languages with an advantage in particular for communication skills."	H	5 3 2	5 3 5	2 3 2
42	*Counselors for a day* Students pretend to be counselors of their local district for a day, to become aware of an active citizenship and communicate it to others.	"The project is valid and has given rise to unexpected inclusions. […] The tasks were divided between teachers and students according to their skills. The kids evaluated the experience well because they felt like protagonists."	H	5 5 3	5 5 3	5 5 3
43	*From the very small to the very large* This work is a light and poetic presentation of two great Italian scientists who, thanks to their work, have given great prestige to their nation.	"… the derivations from Latin of some scientific words have been studied; we synthetized content to write the scripts. In the hours of Natural Sciences, the concepts inherent to the smallest parts of matter, atoms, molecules, were developed and deepened, to then move on to the study of planets, galaxies and black holes. In the Communication Techniques discipline, verbal and non-verbal communication was studied…"	H	5 4 5	5 4 5	5 4 5
44	*Ecojournalists or "Bedroomdesk Activists"* The project aims at helping students to understand that each of us belongs to a larger ecosystem, and that we must be responsible toward our environment.	"Students were not only asked to play an active role in the decision making process, but they felt free to express their own opinions, using creative technological media."	H	5 5 4	5 5 4	5 5 4
45	*Here I am!* A story about education as the key to the fulfillment of the students' most colorful ambitions and most obstinate aspirations, as the missing piece of the puzzle to ensure that the picture of life is complete.	"Taking paper, pen and aspirations, they wrote bold words which they then attached to their backpacks. Words like 'dream, future, ambition, school, study, youth, world, life, curiosity, passion, truth, thought, word, head, heart.' They wrote them in Greek and Latin to show that besides their dreams, they know declinations well. Then they translated them into English, because they love to feel global. And if they could, they would also have included all the other languages of the world. Because it's the world they are	H	5 3 5	5 3 5	5 3 5

		aiming at."				
46	*Gambling: an experience between history and probability of success* The project describes the most famous games of chance, highlighting their evolution over time and the low chance of winning considerable sums.	"From the point of view of competences, attention was paid to the ability to communicate, collaborate and participate both within the class group and with the teacher."	H	5 4 4	5 4 4	5 4 4
47	*The advertising language between history and modernity. From Aristotle to coca cola* A story that, starting from the ancient philosophers, passing through the literature of the '300 and arriving to the modern advertising messages, teaches "to look, not just to see."	"I decided to participate to the project with my students to activate more active and profitable learning processes, which are generally experienced solely as functional to passing a test."	H	5 4 4	5 4 4	5 4 4
48	*Double interview* The story stages impossible interviews, as a couple, with famous people from the scientific world.	"The teacher's goals were: - have the students investigate the topic not only from the purely scientific point of view but also literary and philosophical - promote the ability to grasp links between historically distant events - Stimulate collaboration among students by creating heterogeneous groups, thus stimulating the less performing students."	H	4 3 5	4 3 5	4 3 5
49	*Latinae historiae* Latin comics in which the students identify with the children of that world who lived the situations of an ordinary day and spoke the language of their time.	"With this study method, the pupils were constantly stimulated to think directly in Latin and so they wrote short comic stories in that language about life in ancient Rome. These same stories, drawn on the billboard, were then photographed and inserted in 1001stories accompanied by audio. Latin immediately came alive, in colors, images and sounds."	H	5 4 4	5 4 4	5 4 4
50	*Epidemics in History* The project explores at an interdisciplinary level what is one of the most important events for our generation: the Coronavirus pandemic.	"The goal of this short documentary is to turn culture into an 'active heritage', a useful and extremely precious tool for understanding current events."	H	5 4 5	5 4 5	4 4 5
51	*In our school, 130 years ago, "rose sickness" was treated* History of the students' school building, which in the past was a provincial hospital specializing in the treatment of a disease, pellagra, very common among the farmers of the area.	"In general, this experience offered students the opportunity to reflect on the mixture of ancient and modern that characterizes the area in which their school is located."	H	5 5 5	5 5 5	5 4 5
52	*It is no longer tomorrow* The macro-theme is that of the environment, with a two-fold focus: on one side, the kids' own territory, through the analysis of the local environmental changes, on the other, the global issues that, although apparently distant, are actually particularly close to	"The peculiarity of the work lies in its articulation around a narrative nucleus: we have in fact inserted the problems within a real story, in which the protagonist, a scientist from a recent past, interacts with other characters."	H	5 5 4	5 5 4	4 5 4

	our daily life.					
53	*Science: a "feminine" substantive* Historical excursus on the presence of women in the field of science since ancient times.	"Students worked in groups, each in charge of different bibliographic searches (mainly on the internet)."	H	5 4 5	5 4 5	2 4 5
54	*Cubed Stories: a Doodle is to blame...* During the pandemic, an activity was born that involved history, Italian and biology at the same time. Urbani, Röntgen who were they? Students created 3D biographical cards of scientists whose discoveries revolutionized the history of medicine.	"Each student / student pair produced their own 'cubed story', which they exhibited during streaming meetings by introducing their scientist, answering their peers' questions, and interacting with other peer–scientists in a 'round table' simulation."	H	5 4 5	5 4 5	3 4 5
55	*A monument to Giustino Fortunato* A project on a historical figure who studied the problems of the social and economic crisis in the South of Italy after the national unity.	"The students were initially a little fearful, but starting to work and understand the mechanism of the project, they got excited, especially in the phase of creating the audio and inserting the contents into the 1001stories tool."	H	5 4 5	4 4 5	4 4 5
56	*I wish that day was tomorrow* Analysis and interpretation of letters sent during the Great War by soldiers or civilians who participated in some historical events from the front or from their places of origin.	"The activity has undoubtedly contributed to making students more aware not only of the principles that govern the drafting of creative texts, but of the accuracy they require when they are based on historical events reconstructed with the rigor that belongs to the historian."	H	5 4 5	4 4 4	4 4 5
57	*Water, the world power* A research on water, accompanied by videos of experiments carried out at home by students during the pandemic.	"Each group chose how to present their part of the story: via a video, a presentation, a poetic text, the soundtrack for the poetic text. The product of each group was then corrected and adapted by the teachers."	H	4 5 5	2 2 2	3 4 5
58	*Interview with Bernardo Buontalenti* A documentary about a brilliant artist from Grand Ducal Florence, inventor of ice cream.	"13 simulated interviews with the artist [were made]. Eight kids alternated in the interpretation of the character and the others in the role of interviewer. All students were involved in the script-writing, while a smaller group specialized in editing the videos."	H	4 2 4	4 2 4	3 2 3

Example of the final, public, review (synthesizing the opinions and scores by the 3 reviewers) for the story "Ecology of a virus" (number 23 in the table above): "A narrative definitely in step with the times! Determined to deal with an ecological issue, students and teachers are taken by surprise by the pandemic like everyone else and, after an initial discouragement and temptation to give up everything, they decide to change course and to deal, with great proof of resilience, precisely with the hottest issue of the moment. A good research work, an excellent capacity for synthesis and communicative effectiveness give us back a learning object in which the virus itself tells its own "story," through the words of the students. The idea of combining images and texts in the slides that scroll on the screen is very effective in supporting the user's focus on specific topics. The use of multiple voices in the reading gives a good rhythm and makes you want to find out how the story will end, thanks also to a note of lightness that flourishes in the words and drawings of the children, giving us hope in a moment in which it is a more than precious commodity."

Appendix 4

The teachers' reports on the experience

There follows the schema of the reports teachers were asked to fill in when delivering the digital stories.

Report's fields	Directions/prompts on how to fill the field
Story title	Insert here the title of your story
School	Insert here the name of your school; please specify the school grade (Kindergarten, Primary school, Middle school, Highschool)
Teachers and students involved	Which teachers were involved in the implementation of the project? How many were they? How many students were involved?
Short description of the experience	How the idea of participating was born, how the topic was chosen, how the children were motivated, how other colleagues were involved (if applicable), how families were involved (if applicable) ...
Format of the story	Here you have to specify whether your story is "complete" (i.e., entailing chapters and sub-chapters) or "compact" (with chapters only). [The 1001stories authoring tool allows creating linear stories, composed by a sequence of "chapters," or hierarchical stories where there are also sub-chapters]
Relationship with the curriculum	Please clarify whether the work was: - related to a single subject (e.g., English, history…) - multidisciplinary - interdisciplinary - extracurricular
Subjects involved	Please list all the subjects that were involved in the project
Where / when the work was done	In what context did the educational experience take place? (tick all that apply) - at school, during school hours - at school, beyond school hours - at home - on the territory - other Can you please share some details on this?
Learning goals and educational approach	What learning goals had you set? What pedagogical strategies were put into place in order to achieve them?
Tasks and roles	Distribution of tasks and roles among students and between students and teachers in the various phases of the work: were the students divided into groups? Who chose who should work with whom? The teacher or the students themselves? How were the diverse talents and attitudes taken into account? Were all the students asked to try all the activities or did everyone do what she was best at doing? Did students swap roles? Were groups homogeneous or heterogeneous in terms of performance?
Implementation of the experience	Please describe in details how the experience was organized and implemented, from the topic's identification to the content gathering and refinement up to the upload of the various parts int the authoring tool and the final evaluation of the result.
Space, time, tools	Please describe where, when and using what technological tools the work was done [apart from 1001stories, participants may use a number of other tools, like video or image editing tools]. Where did the project take place? How long did it take to work on the project? How much and what work was done in the classroom and how much at home? What tools were used?
Help / support	Were you supported by someone outside your group? (please tick all that apply) - The students' families/relatives - Local authorities - Experts on the subject (e.g., a museum's curator) - Conventional sources (e.g., the local library) - Internet

	- Not applicable Can you please share some details on this? (e.g., please clarify for what specific tasks you were supported)
Distribution and dynamics of "knowledge"	Please describe who learned what and from whom/what source (teachers, students, books, the internet, experts…).
Overall evaluation	Evaluation of the whole project and of the educational experience: what worked well and what did not work? What problems did you meet and how did you manage to solve them? What educational benefits did the students gain? How can the experience be improved?

Open-Ended Tasks Promote Creativity in Minecraft

Yue Fan[1], H. Chad Lane[1*] and Ömer Delialioğlu[2]
[1]Department of Educational Psychology, University of Illinois, Urbana-Champaign, USA // [2]Department of Computer Education and Instructional Technology, Middle East Technical University, Turkey //
yue.fan@liulishuo.com // hclane@illinois.edu // omerd@metu.edu.tr
[*]Corresponding author

ABSTRACT: Due in part to its flexibility and open design, the video game Minecraft has emerged as a popular tool for teaching and learning. Inspired by prior research showing the influence of problem-solving mindsets in physical settings, this study is an effort to understand the extent to which an open-ended task influences subsequent problem-solving behaviors in a virtual environment. Specifically, we investigate creativity and its relationship with task design in Minecraft by comparing a well-defined task group, instructed to follow step-by-step directions, with a group pursuing an open-ended task requiring a higher degree of agency. Creativity is assessed using two conventional approaches: the Alternative Uses Test (AUT) and the Consensual Assessment Technique (CAT). Judges were trained to evaluate using both methods and achieved sufficient agreement on a subset of the data prior to completing the full data set. Our results suggest that (1) participants who engaged in the open-ended task receive significantly higher CAT scores than those in the well-defined task group, and (2) among variables such as the level of skill/experience, interest in Minecraft, and materials (blocks) used in Minecraft, only game interest level has a significant influence on the CAT score.

Keywords: Creativity, Minecraft, Problem solving, Educational games, Computer-based learning environments

1. Introduction

Counter to early research that tended to focus on the potentially negative consequences of playing video games, the last decade has witnessed growing attention on the possible benefits of playing games, including investigations into learning, motivation, engagement, and creativity (Granic, 2014; Connolly et al., 2012). Research on the design and effectiveness of educational games has also grown dramatically suggesting that well-designed games can be effective in enhancing learning-related outcomes (De Freitas, 2018; Mayer, 2019). This body of work has laid a foundation for further study about how specific games and game genres may influence learners and be effectively integrated into educational programs.

Minecraft is one of the most popular video games in history with over 200M copies sold and 131M active monthly users (Watts, 2021). Minecraft is typically classified as a "sandbox" game, which means that players have a very high level of agency and freedom to pursue their own goals and explore/engage in creative activities in whatever ways they choose (Brand & Kinash, 2013). Studies on the impacts of sandbox games, and Minecraft in particular, are only now beginning to emerge. The focus of our work is on how Minecraft-based experiences may be deployed to promote creative thinking. The research reported here seeks to fill a gap in this literature and help educators design game-based learning activities to allow their students to explore their creativity and interests.

Not only is Minecraft popular, it has also been adopted by teachers and educators all over the world (Pusey & Pusey, 2015) due to its deep ties to Science, Technology, Engineering, and Math (STEM) learning (Lane & Yi, 2017; Short, 2012). One of the reasons for these direct links is because Minecraft can be viewed as a scaled-down simulation of the natural world and interactive representation of a wide range of STEM-relevant phenomena. Typical game activities involve exploring a range of different biomes, locating and collecting resources, interacting with animals and sea creatures, building large structures, farming, and much more. Given this rapidly growing interest, there is a pressing need to provide empirical support and evidence-based principles for designing content and leveraging the rich resources provided by Minecraft (Baek et al., 2020). The study reported in this paper focuses directly on a task that could easily serve as a learning activity (building a home), and so it has implications on how to design learning tasks that allow learners to have more freedom and ability to express themselves. Specifically, our findings suggest that less rigid structure is more likely to promote creativity in goal-driven learning activities in Minecraft.

In public dialogue, it is common to hear Minecraft described as a form of "digital Legos." Although intuitive and conceptually accurate since both activities use "blocks" as the standard unit of manipulation, the claim does not genuinely convey the full interactive capabilities of Minecraft. For example, Minecraft simulates a variety of real-world phenomena such as flowing water and plant/tree growth. It also allows for the construction of large,

complex machines with *Redstone*, Minecraft's version of electricity, enabling players to engage in automation of a range of science-relevant tools for farming, exploration, and building.

The study reported here represents a partial investigation of the accuracy of the common metaphor between Minecraft and Legos. Specifically, we report a replication of earlier work with Legos that explored the relationship between task openness and creativity (Moreau & Engeset, 2016). The key finding from this work was that open-ended tasks with Legos (e.g., "build a house") fostered greater creativity than well-defined, more narrowly defined tasks that do not allow significant choices to be made by a learner (e.g., "build a house exactly to specifications"). We sought to determine if these findings would hold up in a digital environment, thus suggesting that creativity emerges in similar ways in both tangible and virtual environments.

As with Moreau and Engeset (2016), we investigate (1) whether there are creativity differences between open-ended tasks and well-defined tasks, i.e., whether a well-designed problem-solving mindset would produce lower creativity scores in a subsequent free-play task. And (2) whether interest, skill/experience of the participants as well as the number of different types of blocks used to build would affect creativity scores of the participants. We pursue the following two research questions:

- To what extent does the use of open-ended tasks in Minecraft influence a player's creativity?
- Does creativity have a significant relationship with level of interest, skill/experience, and the number of different types of blocks used to build (when given the choice to use as many as desired)?

In the remainder of this article, we first describe Minecraft in more detail, arguing that it qualifies as an authentic learning environment then describe our partial replication of the Moreau and Engeset (2016) study. We conclude with a discussion of the findings and implications on practice.

2. Minecraft as an authentic learning environment

Minecraft is fundamentally a world made up of blocks (earning it the Lego analogy) that come in a wide range of types (see Figure 1). These blocks can easily be manipulated (e.g., created, destroyed) and have vastly different properties and functions. Some blocks are solid, harder or softer, transparent or opaque, "sticky", or even a liquid. The name *Minecraft* reveals the nature of playing: typically one collects materials (as blocks) and then combines those in ways to create new tools and resources. In turn, the results unlock new capabilities and enable more advanced forms of play. Players collect resources in the world by *mining* (often with a tool, such as a pickaxe), and then learn to *craft* new items that further enable building of more complex artifacts, such as bridges, buildings, machines, and farms.

Figure 1. A selection of blocks available in Minecraft each with unique properties and conditions for use

To customize the game, Minecraft can be played (1) either by individuals or in groups, (2) locally or on a server, and (3) with the use of "mods" (modifications to the basic game). Players can play in *survival* mode, which requires resource collection, some combat, and the creation of structures to protect oneself from enemies. In *creative* mode, on the other hand, the player has an unlimited supply of resources, no need for food, no chance of injury, and the ability to fly freely around the world. In the study reported below, a hybrid mode was used that

enabled free play, but with no combat or chance of taking damage. In our study, a limited set of blocks were provided to best estimate the Lego study in which a fixed set of LEGOs were provided to participants prior to beginning the task (Moreau & Engeset, 2016).

Minecraft provides practically unlimited opportunities to manipulate and customize their in-game experience. Intuitively, it is not difficult to link the openness of this and other sandbox games to potential benefits in creative thinking and self-determination. By definition, sandbox games promote intrinsic motivation by providing a context for self-generated goals and natural autonomy (Tichon & Tornqvist, 2016). One of the most prominent features of sandbox games are their openness and the ability to explore vast computer-generated spaces however a player desires (Brand & Kinash, 2013). Minecraft exemplifies these freedoms: exploring the diverse landscapes and biomes, crafting various new blocks, customizing one's living space, collaborating and coordinating with friends, "modding" to add unique features, and more. This extreme level of choice provides an ideal context for creative problem solving and expression.

Minecraft can also be viewed as an *authentic learning environment* in the sense that it shares many of the same characteristics of the natural world (e.g., weather, terrain, different biomes, living creatures, caves, mountains, flowing water) and enables many of the same activities that humans engage in (e.g., exploring, building, collaboration, resource collection, construction of machines with moving parts). Considering Herrington & Oliver's (2000) framework for authentic learning environments, Minecraft faithfully reproduces many real-world settings, provides opportunities to engage in realistic activities (e.g., building complex structures), provides easy access to expert content (e.g., through mods or content built in world), the ability to take on multiple roles (e.g., surveyor, designer, or builder), and includes direct support for collaboration through a shared workspace and chat (p. 4-6). The claim that Minecraft includes sufficient complexity to be considered a realistic learning environment is also supported by its growing use as a testbed for real-world problems. For example, Project MALMO at Microsoft uses Minecraft for the development of machine learning models in robotics (Johnson et al., 2016). In addition, Facebook created *CraftAssist,* a tool for studying human-agent collaboration around complex joint tasks and capture of natural language interactions (Jayannavar et al., 2020). In sum, Minecraft, while not originally designed as a simulation tool, provides a wide range of features and capabilities that enable it to recreate problem solving contexts that can resemble key aspects of the natural world.

3. Creativity, Legos, and Minecraft

In this section, we clarify our definition of creativity and provide background necessary for our study.

3.1. Big C vs. Little c Creativity

Research on creativity tends to address two broad categories of creativity: Big "C" (sociocultural definition) and little "c" (individualist definition) (Sawyer, 2012). Big "C" refers to "the generation of a product that is judged to be novel and also to be appropriate, useful, or valuable by a suitably knowledgeable social group" (Sawyer, 2012, p.8). This means those products with widespread social approval, solutions to societal problems, or significant works generated by famous artists or musicians are frequently cited as examples. For example, the Mona Lisa painted by Leonardo da Vinci and universal gravitation as articulated by Newton both qualify.

Little "c," on the other hand, is defined as producing "a new mental combination that is expressed in the world" (Sawyer, 2012, p. 7). In this case, creativity is not limited to validation by a large group of people, but extends to all people who are solving everyday problems. Everyone exhibits little "c" creativity to some capacity since daily life involves the combination of elements to create new products or to solve simple problems in a novel way. There is no restriction that others have not used the same solution. For example, someone who has toothpaste, but no toothbrush might brush their teeth with their finger, it achieves a similar goal and is a creative solution. Although we focus on little "c" creativity in our study, one can certainly not rule out the potential for big "C" creativity to be possible in Minecraft. Indeed, massive undertakings are not uncommon in the game that take years of effort, specialized skills, and collaboration (Peckham, 2013).

3.2. Divergent thinking and convergent thinking

Research has identified at least two ways of thinking that contribute to creativity: divergent and convergent thinking. Divergent thinking expands a problem solution space and involves experimentation to identify and

develop multiple ideas, each of which could possibly become a solution. In contrast, convergent thinking prunes a problem-solving space and "emphasizes speed, accuracy, and logic" in pursuit of "the single best (or correct) answer to a clearly defined question" (Cropley, 2006). Creativity can be regarded as a cyclic process of ideation (i.e., the formation of ideas) that relies on divergent thinking to expand a set of potential solutions followed by evaluation of those ideas to identify the best options, which relies on convergent thinking (Lubart, 2001).

Minecraft can be considered as a tool that inherently promotes divergent thinking for players given its openness emphasis on building, and high degree of collaboration. During regular game play, players repeatedly synthesize information to make decisions and implement different approaches for building and coordinating activities. It is simple (and natural) to look at projects from multiple dimensions (top-down, inside, outside, etc.) and in different natural settings. Players fully explore the virtual world with a broad range of possible interactions in Minecraft, employ trial and error, and learn from failure (Green & Kaufman, 2015), which suggests divergent thinking is implicitly encouraged. The virtually limitless sets of combinations of virtual blocks provide further fuel for the argument. Lastowka (2011) explains that "Minecraft is rooted in the free exchange of creativity and users' creations." In creative mode, players construct buildings, tools, machines solely for the purpose of creative expression (Garrelts, 2014).

Research has shown that creativity can be boosted during group work and via interactions between creators (Sawyer, 2012). This is similarly compatible with Minecraft since playing collaboratively in a multiplayer mode is one of the most common ways to play. By communicating with others, people engage in different ways of thinking, including use of imagination, perception, and reasoning. Such collective capabilities are crucial to creativity (Shabalina et al., 2015). Note that while the study reported here did not investigate collaborative creativity, the collaborative affordances of Minecraft suggest important future directions of research.

3.3. Creativity and the problem-solving mindset

In the context of using creativity to solve problems from daily life, the *problem-solving mindset* has a close relationship with our creativity. According to research on problem-solving mindsets, cognitive activities that relate to problem solving will sustain over time and contexts (Moreau & Engeset, 2016). The problem-solving mindset refers to a phenomenon that people's behavior or thinking processes in one case can influence their ideas and performances in later, unrelated tasks. Since the formation of mindset is largely affected by the problem space in which the learner works, it is possible to manipulate perceptions of a problem space and see how those differences influence creativity.

Research has revealed three key components that are relevant for understanding a problem space: (1) the initial state (problem itself), (2) the set of operators (rules and strategies) that enable one to proceed from the initial state to (3) a goal state (the solution) (Davidson & Sternberg, 2003). According to Kitchener (1983, p. 223) *well-defined* problems are those "for which there are absolutely correct and knowable solutions," while *ill-defined* problems are those "for which there are conflicting assumptions, evidence, and opinion which may lead to different solutions." Well-defined and ill-defined problem spaces can be loosely thought of as aligning with convergent thinking and divergent thinking. Divergent thinking is most closely associated with open-ended problems where solutions are open to interpretation and may have unclear/unspecified operators. Convergent thinking, on the other hand, is more closely related to well-defined problems in the sense that a clear and specific initial state is given and an operational path to reach a correct answer exists in the problem space. Based on connection between problem solving and creativity theory, experiments can be run to set up which kind of mindset best promotes learner creativity and their ability to solve problems creatively.

3.4. How Minecraft and Legos are believed to promote Creativity

Substantial evidence has emerged supporting the connection between creativity and game design (Green & Kaufman, 2015). For example, in a study of almost 500 children, videogame play was found to be related to multiple dimensions of creativity, whereas other information technology use was not (Jackson et al., 2012). In terms of supporting creativity, Melián Díaz et al. (2020) found increases in divergent creativity after presenting teachers in training a Minecraft task to "build the home of your dreams" using a different measure of creativity than the study presented here. Preliminary research has also shown that the openness and basic play of Minecraft has positive effects on learner creativity. "To play Minecraft is to use the game as a creative tool. Minecraft requires players to be creative, even if that creativity is limited to designing a crude shelter or tunneling the layout of a mine" (Lastowka, 2011). Despite these promising preliminary findings, we are aware of no rigorous research demonstrating causal links between Minecraft play and creativity. With its high level of customization,

block types, and specialized tools, Minecraft enables high levels of expressivity which suggests it is an ideal environment for promoting creative thinking. Further, the openness of Minecraft allows for experimentation with various conceptions of narrative and characterizations of plot lines, further aligning it with conditions that known to promote creative production (Cipollone et al., 2014).

The Lego and creativity studies conducted by Moreau and Engeset (2016) were the inspiration for our study. This research used Legos and sought to confirm the notion of problem-solving mindsets described in section 3.3. Researchers decomposed the problem-solving space to the initial state, operator and goal state. The degree of openness was manipulated in order to investigate the influence of activated problem-solving mindsets on subsequent creative tasks. Three experiments are reported in their article, with the first experiment forming the basis for our Minecraft replication. This study compared three different mindsets (well-defined vs. open-ended vs. control) and their influence on a subsequent task (open-ended vs. well-defined). The findings showed that a well-defined mindset (instructions with clear goal) diminishes performance on a subsequent open-ended task (no instruction, or goal). Our study follows the same design.

4. Methods

4.1. Research design and variables

Participants were randomly placed into the experimental or comparison group upon arrival. The independent variable was the task design (open-ended or well-defined problem-solving task) while dependent variables were (1) a conventional creativity measure, the Alternative Uses Test (AUT) and (2) a free-play building task in Minecraft for use with the Consensual Assessment Technique (CAT). Here, the experimental procedure was simplified by ignoring the measurements that have a distant relationship to creativity, such as enjoyment which was included in the Lego study (Moreau & Engeset, 2016). Our experiment focused on the relationship between problem-solving space (initial state, operators, goal state) and subsequent task performance. Further, participants were given 15 minutes in their free play (second task), which was the product assessed in the CAT (dependent measure of creative expression). To best mirror the Lego study, participants were given a fixed set of Minecraft blocks from which draw in this free-play task. In summary, we used AUT assessment as in the Lego study (Moreau & Engeset, 2016) with an additional measurement added, the CAT, to capture creativity in different ways.

4.2. Participants and characteristics

The participants of the study were 42 undergraduate students from a university in the midwestern section of the United States who fit the requirement of having basic experience with Minecraft (i.e., they played Minecraft before and at least knew basic operation of playing the game). We administered a survey to gather basic demographic information and to capture the Minecraft playing characteristics of the participants. Among those participants, 74% of participants were male and 26% were female. Two main categories of questions were included in the survey (1) the skill/experience level with video games generally as well as specific Minecraft experience, and (2) the interest level and motivation to play Minecraft. Unsurprisingly, skill/experience level had a high correlation to interest level ($r = .70$), meaning simply that people who liked the game played it more often.

Table 1. Participant familiarity with common Minecraft activities

Activity	Rating
Collecting/mining resources	4.69
Farming (planning, harvesting crops)	4.31
Crafting tools; using crafting tables	4.29
Planning, designing, and building	4.00
Redstone (i.e., electricity, machines)	2.64
Command blocks	2.36
Creating/maintaining a Minecraft server	2.26

Note. Rating ranges from 1 (low familiarity) to 5 (high familiarity).

All participants had some prior exposure to Minecraft which ranged from over 10 years of experience to less than a month. 38% still played at least once a month and the remaining 62% had not played recently (within a month). Table 1 shows participant ratings of their experience with several common Minecraft activities, revealing high familiarity with many (mining, farming, building) but less exposure to advanced play (Redstone,

command blocks). Only basic skills were required to fully participate in the study. Finally, self-reports on the appeal of Minecraft using a scale of 1-10 (10 = high), 19% gave a score under 6 points, 52% liked Minecraft (7-8 points), 28% expressed high interest with a rating of 9 or 10. In sum, these data indicate participants in the experiment exhibited sufficient knowledge of Minecraft and most found it to be an enjoyable way to spend time.

4.3. Procedure

The research procedure that participants were guided through took roughly 40 minutes to complete (Figure 2). Upon arriving, participants completed a 4-minute survey, which included their demographic information, and their interest in and experience with Minecraft (results presented in section 4.2). Participants were then situated in the Minecraft environment with a basic landscape and provided with a reminder of how to play if needed. They were then invited to move around in-game and acquaint themselves with workspace for a few minutes. Participants then completed a 15-minute problem-solving task in Minecraft that was either well-defined (group 1) or open-ended (group 2), assigned randomly. All participants were provided a collection of resources from which to build in the game (in chests). The well-defined task group was given printed instructions and a clear direction to "follow the step-by step instruction, build the house with the bricks in the chest" with pictures. The resources provided precisely what was needed to construct the sample house (see Figure 3).

Figure 2. Procedure of the Minecraft creativity study

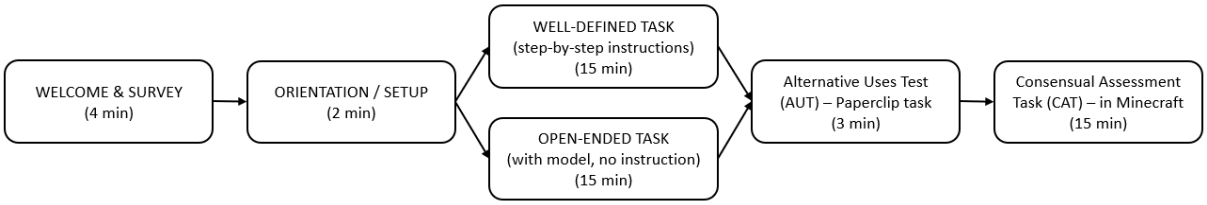


Figure 3. Model house used in the well-defined task group

Front Back

Front door Side face

Participants in the open-ended condition were also given a picture of a house, but with the general directive to "build a house like the sample, with the blocks of your own choice." A chest with varied materials (64 kinds of blocks) were provided for the participants to build in a wide range of styles and to a larger scale. No further guidance was provided.

Upon completion of the task, all participants first completed the Alternative Uses Test (AUT) that asked for as many uses of a paperclip can be generated in 3 minutes. Finally, participants were invited to return to Minecraft

for a 15-minute session of free play in Minecraft with instructions to build anything they wanted in full creative mode (no limit on blocks that were made available). The product of this session was used for our Consensual Assessment Test (CAT). Screenshots and saved game maps of the product they created in Minecraft was collected for CAT scoring.

4.4. Data collection instruments and analysis

The initial survey consisted of 10 questions covering their frequency, history, preferences, motivation, skills, as well as their history with Minecraft. The results of the survey were summarized in section 4.2.

The Alternative Uses Test (AUT) (Guilford, 1967) and Consensual Assessment Technique (CAT) (Amabile, 1982) are two of the most common choices for judging levels of creativity. The AUT assessment seeks to capture divergent thinking and has repeatedly been shown to correlate with additional measures of creativity (Hocevar & Bachelor, 1989). The CAT assessment seeks to measure creativity as expressed through a product or artifact and is considered one of the best measurements of creativity in terms of reliability, discriminant validity, and nomological validity. Multiple experts are needed to rate the creativity of a collection of products by comparing them with one another individually and without outside guidance (Amabile, 1982). This test focuses on little "c" creativity and builds on the intuitive theory that the combined assessment of experts in certain field is the best measure of the creativity of a product. The product of a CAT is usually scored by at least three experts who use their own professional sense of what is creative in a domain (Kaufman & Baer, 2012). The testing objects used in prior CAT assessments have covered a wide range of artifacts, including engineering and artistic creations (Baer & McKool, 2009), which make it suitable for assessing open-play Minecraft products.

We used three expert Minecraft players as judges who were not involved in the research project. They evaluated the 42 participants' free-play Minecraft products (i.e., "build anything") without awareness of which condition a participant was assigned to. The overall inter-rater reliability was acceptable between the three (irr = 0.52), with the correlation between the three raters (R1, R2 = 0.57; R1, R3 = 0.50; R2, R3 = 0.40) similarly acceptable based on the subjective nature of CAT grading. A *t*-test was then used to assess the effects of well-defined task and open-ended task on participants' CAT mean score among three judges.

The AUT asks participants to generate as many uses as possible for a common household object. In this study, participants were asked to "list the different uses of a paperclip as much as possible in three minutes." We used two of Guilford's (1967) dimensions that most directly reflect divergent creativity: (1) *originality*, which refers to a statistically uncommon answer when compared to the overall data set, and (2) *fluency*, indicated by the quantity of appropriate responses given. Three independent raters judged whether each answer was appropriate (reasonable), unusual (5% of the group mentioned) or unique (1% of the group mentioned). Following the AUT official manual (Wilson et al., 1954), we required at least 2 of the 3 judges to indicate that an answer was appropriate in order to include it in the measurement.

To complete data analysis, two judges independently completed their AUT scoring, with a third judge used to break ties when they arose. To derive a score, participants received one point for each appropriate response (fluency). From all appropriate responses, then, 1 point was awarded for each "unusual" answer and 2 points for each "unique" answer. For example, if a participant generated 15 answers to the AUT assessment and 10 of them are judged as appropriate, they are assigned 10 points for fluency. If among those 10 answers, only 2 of them are "unusual" and 1 "unique." If those two answers that are mentioned by less than three people (5% of the group) and one response mentioned only once among 48 participants, then the participant could earn $2\times1+2=4$ points for originality. The AUT score for this participant is to add the fluency and originality scores to get 14.

Finally, for the free play "build anything" task in Minecraft, the result was graded via the CAT assessment by experts. Three experts made their judgments about the product creativity with score (range: 1-10), independently based on their own criterion (no standardized judging criterion, a common design choice in creativity research).

5. Results

5.1. RQ1: The relation between task type and creativity

To test if there was a significant difference on creativity score between the two kinds of problem-solving tasks discussed above, the creativity score of both groups using the CAT and AUT were considered. For the CAT

assessment, scores were found to be normally distributed, thus meeting the requirement for a two-sample unequal variances *t*-test (see Table 2). AUT scores with 42 participants did not fit the normal distribution ($p = .04 < .05$ in Shapiro-Wilk normality test), meaning hypothesis verification for the two groups required a non-parametric test. For this, we used Mann-Whitney U test basing it on the continuity correction of the rank in group.

Table 2. CAT scores for well-defined vs. open-ended tasks

Task	*M*	*SD*	*n*	*t*-test	*df*	*p*
Well-defined task	5.06	1.58	22	2.14*	40	.038*
Open-ended task	6.07	1.42	20			

Note. *$p < .05$.

Since participants in the well-defined condition received a lower creativity score ($M = 5.06$) than those in the open-ended condition ($M = 6.05$; $t = -2.14$, $p = .038 < .05$, Cohen's $d = 0.66$), the main effect of the problem-solving task on CAT score was significant. The open-ended task produced significantly higher CAT scores than the well-defined task in the final "build anything" free-play task.

The AUT × (well-defined vs. open-ended task), Mann-Whitney U test suggested no effect was detected. The well-defined group scored higher ($M = 10.68$) than the open-ended task ($M = 8.55$), with $U = 161$, $W = 371$, although this was not significant ($p = 0.136$). Thus, there is no significant difference between well-defined mindset and the open-ended mindset in their AUT performance of the following free-play problem-solving task.

5.2. RQ2: The relationship of level of interest, skill/experience of the participants, and the number of different types of materials used with creativity

Apart from creativity difference between the well-defined problem-solving task and the open-ended task in Minecraft, we consider three variables that might have effects on the creativity score: (1) skill/experience level with Minecraft, (2) interest level in the game, and (3) the different types of blocks used in the open task. All the data were quantified and divided into high and low level in two categories: skill/experience and interest in Minecraft. To test if there are any relationship between the variables and their creativity score, a *t*-test was used with three groups of comparison: CAT × (low skill/experience level vs. high skill/experience level), CAT × (low interest level vs. high interest level), and CAT × (low block use vs. high block use).

Table 3. CAT scores for skill in Minecraft, interest in Minecraft, and for overall number of different blocks used.

	Low level			High level			*t*	*p*	*df*
	M	*SD*	*n*	*M*	*SD*	*n*			
Skill	4.91	1.87	17	5.95	1.18	25	2.03	.051	25
Interest	5.08	1.63	22	6.02	1.38	20	-2.01*	.049*	40
Blocks	5.21	1.69	19	5.79	1.45	23	1.19	.239	36

Note. *$p < .05$. The CAT (Consensual Assessment Technique) is a subjective judgment by domain experts on the level of creativity present in an artifact.

This analysis (Table 3) showed first, that those who have low level of skill/experience had a lower CAT score ($M = 4.91$) than those who are of high level of skill/experience ($M = 5.95$; $t = -2.03$, $p = .051$, marginally significant). For the test of interest level, the high-interest group (M = 6.02) had a significantly higher CAT score than the low-interest group ($M = 5.08$, $t = -2.01$, $p = .049$), with Cohen's $d = 0.62$ (medium). Lastly, in terms of the number of different blocks used in free play, no significant difference was found between those with a high level of block use ($M = 5.79$, $t = -1.19$, $p = .24 > .05$) than those with low counts ($M = 5.21$).

Correspondingly, we did the same t-test for each of the variables using the AUT scores (Table 4). Neither skill/experience ($M_{low} = 9.47$, $M_{high} = 9.80$, $t = -0.23$, $p = .82 > .05$) nor blocks used ($M_{low} = 9.68$, $M_{high} = 9.65$, $t = 0.02$, $p = .98 > .05$) were statistically significant. However, interest level (in Minecraft) was significant for the AUT test ($M_{low} = 8.27$, $M_{high} = 11.2$, $t = -2.10$, $p = .04 < .05$) with a medium effect size (Cohen's $d = 0.65$). Thus, participants who have high level of interest on Minecraft also had significantly higher AUT scores.

To summarize, when comparing the CAT and AUT assessment together to see if variables other than task condition had influence on participants' creativity, neither skill/experience level nor number of different blocks used had significant influence. However, a high level of interest in Minecraft is associated with higher AUT and CAT scores, perhaps suggesting that more creative people are naturally drawn to Minecraft to begin with.

Table 4. AUT scores for skill in Minecraft, interest in Minecraft, and for overall number of different blocks used

	Low level			High level			*t*	*p*	*df*
	M	*SD*	*n*	*M*	*SD*	*n*			
Skill	9.47	4.09	17	9.80	5.11	25	-0.23	.818	39
Interest	8.27	4.17	22	11.20	4.82	20	-2.10*	.042*	38
Material	9.68	4.27	19	9.65	5.08	23	0.02	.982	40

Note. $^{*}p < .05$. The AUT (Alternative Uses Test) is a common assessment of creative thinking that asks participants to imagine as many different applications or uses of an object that they can generate (e.g., a paperclip in our case).

6. Discussion

Our analysis of CAT (Consensual Assessment Technique) scores, used to assess Minecraft products created during a subsequent task, produced a statistically significant difference in creativity present in favor of the open-ended condition. The significantly higher CAT scores for the open-ended task supports the idea that open-ended tasks promote creativity in Minecraft more so than well-defined tasks do. However, this was not replicated for the external creativity test, the AUT, where we found no significant difference between the two conditions.

One possible reason is that we chose to use of a simpler assessment of divergent thinking than the original Lego study. In particular, the significant result from Moreau and Engeset's (2016) Lego study 1 is measured by a version of Torrence Test of Creative Thinking (Kim, 2006), a more complex assessment than the one used in our study. Lego study 2 was designed to compare the creativity difference between instruction (present vs. absent) and outcome (present vs. absent). In this study, the researchers found a higher AUT score resulted from well-defined mindset than the open-ended mindset, which was not contradicted in our study (our well-defined group did have a higher mean AUT score, but it was not significant). Another potential explanation is that the AUT is simply too far of a transfer test as it has nothing to do with Minecraft, Legos, or construction activities. The results suggest many avenues for future research, such as using the Torrence test after Minecraft play, or even modifying the AUT to focus on Minecraft-related content (e.g., many Minecraft objects can be used in novel, unintended ways such as using pressure plates as shelves, or stair pieces as chairs at a table).

Based on the comparison between Lego and Minecraft studies related to research question 1 (Do open-ended tasks promote creativity?), we can conclude that the key result from the Lego study (Moreau & Engeset, 2016) is comparable and that the finding holds up in Minecraft. This suggests that considering Minecraft a virtual form of Legos is a reasonable metaphor. This is perhaps not surprising due to the fundamental similarities between the two experiences – both are block-based and present very similar interaction mechanics. If anything, Minecraft may have potential to promote creativity in ways that go beyond Legos simply because of additional features that are present in Minecraft but not Legos. For example, Redstone enables the construction of machines, switches, lights, and moving parts. Further, in Minecraft it is possible in incorporate simulated plants and animals, and work in the context of a simulated natural world (e.g., with mountains, streams, lakes, and more). Of course, the fundamental strength of Legos is that it is physical, which has been consistently been shown to have a variety of benefits over purely virtual interfaces (Schneider et al., 2010). Moving forward, it will be important to investigate the ways in which physicality can be leveraged and hybrid learning environments can be used to reap benefits from both. It is promising, for example, to see Minecraft being used to allow learners to explore the use of 3D printers in design tasks (Niemeyer & Gerber, 2015).

In terms of theoretical perspectives on the study, we view our results as supportive of the important role of the design of problem-solving tasks in creativity research and education. Generally, the intervention likely qualifies as a version of *creativity priming*, in that it seeks to establish cognitive and affective conditions that influence future behavior (Sassenberg et al., 2017). Additionally, we found that prior interest in Minecraft contributed to a higher level of creativity present in both the AUT and CAT assessments. Prior research has demonstrated consistent associations between advanced interest and the accumulation of knowledge (Hidi & Renninger, 2006). If one infers from our survey results that greater interest in Minecraft is associated with greater knowledge of the game, our study further supports the established idea that creativity is bolstered by prior knowledge (e.g., Ward, 1994). Specifically, knowledge of and experience with the game of Minecraft seemed to enable greater levels of creative expression in our subsequent open-ended task.

Creativity has been a consistent focus from researchers for over 50 years (Sawyer, 2012), with much of this work looking at how to enhance creative thinking. Our work investigated the role of a priming task (well-defined or open-ended) and its impact on subsequent creative activities. In this study, two research questions were raised

that were inspired by a prior Lego study (Moreau & Engeset, 2016). The intent was to investigate whether and to what extent, the findings would hold up in Minecraft – that is, do open-ended tasks produce higher levels of creativity in future tasks? If so, it would lend credence to the often-made suggestion that Minecraft is a "digital form" of Legos. Our findings suggest that in-game creativity was enhanced, however that this creativity did not extend to a general, non-Minecraft activity. Further research should investigate both the duration and nature of this finding, and look more deeply into how to design learning-related tasks in the game. Given Minecraft's continued growing popularity and use in education, we anticipate a consistent and growing need for further empirical research with the game.

6.1. Implications and future directions of research

The results of this study could be relevant to how instructors choose to use Minecraft. When it is possible to provide less structure in a task, using less specific guidelines and rules may have downstream benefits. Learners might enjoy it more (as was the case in the Lego study) and there may be additional opportunities to leverage creative thinking in subsequent tasks as learners grow more accustomed to receiving less guidance. Also, as discussed, Minecraft provides many STEM-relevant opportunities for learning and thus the chance to explore creativity in a STEM context could be highly appealing to educators. The potential to increase its authenticity and relate it to real-world phenomena needs continued investigation. For example, researchers are now using Minecraft as a tool for engaging children in Urban Planning and to communicate their vision for the cities they live in (Andrade et al., 2020). Combining architectural and civil engineering goals in a Minecraft-based curriculum could set seeds for a new generation of creative cities and innovative solutions to sustainability and environmental challenges.

Potential links between aptitude and creativity are worthy of further research – learners with lower aptitude tend to benefit more from closed tasks, whereas high aptitude learners are comfortable with open-ended learning challenges. Unpacking the relationship between these two, learning, and creativity could have benefits. Further, investigation of learning paths that gradually reduce guidance and scaffolding (Reiser & Tabak, 2014) such that open-endedness is presented gradually, and perhaps made available to learners who may never choose such tasks without such support.

Another possible avenue is to extend the comparison between Lego and Minecraft and their effects on creativity. It would be valuable to pursue mutual and cumulative impacts on each and see if there any potential significant differences between their influence on creativity. For example, a Lego-based approach may be specialized in a tangible interaction (Horn & Jacob, 2007), which makes it more easily to transfer the knowledge and be accepted by students. While Minecraft may have the advantage of constructions with more interactivity and of much greater scale. Advanced functions in Minecraft may deepen links to STEM fields, such as with Redstone and "Command" blocks, which link coding/computing into the game directly.

6.2. Limitations

The current study has several limitations that should be considered. The first is the acceptable but less than desired inter-rater reliability for the CAT assessment (kappa = 0.52). This may have been a result of how the data was captured (screenshots of the game), or our decision to not pursue a common metric for judges for evaluating the products of the participants. Also, based on the structure of Intellect from Guilford and Hoepfner (1971), divergent thinking levels of participants are closely associated with their intelligent level, which was not accounted for in our study. Thus, the result could be selective in the group of participants in this study. Another important limitation of our intervention is a short duration (40 minutes) bringing the generalizability and stability of our results into question (we note, however, that our model study from Moreau and Engeset (2016) consisted of roughly the same length of time). Finally, we note that we chose to not use a pre-post test design, which limits the ability for the study to suggest a causal connection between condition and creativity scores. It is indeed possible that creative people simply score highly in both environments. It is a methodological challenge to do a creativity post-test due to inherent testing effects (that is, the test itself may activate creative thinking). A more complicated future study could perhaps give creativity pre-tests far in advance to the intervention to minimize this kind of effect.

7. Conclusion

From programming, to sales, to STEM careers of all kinds, the ability solve problems creatively is seen as critical for success in the modern world (Sawyer, 2012). The question of how to design learning interventions in ways that simultaneously engage learners and promote creative thinking is critical in order to meet this growing demand. Our work addresses the tension that arises between well-defined tasks versus those that are more open-ended and suggests that providing opportunities with less guidance and more loosely structured learning activities may have downstream benefits in terms of promoting creative thinking. This study showed a basic replication in Minecraft of prior work using Legos suggesting that a digital environment may be similarly as effective to promote creativity. It also showed that prior interest in Minecraft contributed to heightened levels of creativity in a subsequent task. Future work should investigate leveraging the power of both virtual and physical environments as well as the design of more flexible plans for learners with variable levels of comfort with open-ended work.

Acknowledgement

Ömer Delialioğlu was supported by the Fulbright Visiting Scholar Program by the Turkish Fulbright Commission for the enrolment in this research.

References

Amabile, T. M. (1982). Social psychology of creativity: A Consensual assessment technique. *Journal of personality and social psychology, 43*(5), 997-1013. https://doi.org/10.1037/0022-3514.43.5.997

Andrade, B. D., Poplin, A., & Sena, Í. S. D. (2020). Minecraft as a tool for engaging children in urban planning: A Case study in Tirol Town, Brazil. *ISPRS International Journal of Geo-Information*, *9*(3), 170. https://doi.org/10.3390/ijgi9030170

Baek, Y., Min, E., & Yun, S. (2020). Mining educational implications of Minecraft. *Computers in the Schools, 37*(1), 1-16.

Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education* (pp. 65-77). IGI Global.

Brand, J., & Kinash, S. (2013). Crafting minds in Minecraft. *Education Technology Solutions*, 55, 56-58.

Cipollone, M., Schifter, C. C., & Moffat, R. A. (2014). Minecraft as a creative tool: A Case study. *International Journal of Game-Based Learning*, 4(2), 1-14.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A Systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661-686.

Cropley, A. (2006). In Praise of convergent thinking. *Creativity Research Journal*, *18*(3), 391–404.

Davidson, J. E., & Sternberg, R. J. (2003). *The Psychology of problem solving*. Cambridge University Press.

De Freitas, S. (2018). Are games effective learning tools? A Review of educational games. *Educational Technology & Society*, *21*(2), 74-84.

Garrelts, N. (2014). *Understanding Minecraft: essays on play, community and possibilities.* McFarland & Company.

Granic, I., Lobel. A., & Rutger, R. (2014). The Benefits of playing video games. *American Psychologist*, *69*(1), 66-78.

Green, G. P., & Kaufman, J. C. (2015). *Video games and creativity*. Elsevier.

Guilford, J. P. (1958). Can creativity be developed? *Art education, 11*(6), 3-18.

Guilford, J. P. (1967). *The nature of human intelligence.* McGraw-Hill.

Guilford, J. P., & Hoepfner, R. (1971). *The Analysis of intelligence*. McGraw-Hill.

Herrington, J. & Oliver, R. (2000) An Instructional design framework for authentic learning environments. *Educational Technology Research and Development*, *48*(3), 23-48.

Hidi, S., & Renninger, K. A. (2006). The Four-phase model of interest development. *Educational psychologist*, *41*(2), 111-127.

Hocevar, D., & Bachelor, P. (1989). A Taxonomy and critique of measurements used in the study of creativity. In *Handbook of Creativity* (pp. 53-75). Cambridge University Press.

Horn, M. S., & Jacob, R. J. (2007). Designing tangible programming languages for classroom use. *Proceedings of the 1st International Conference on Tangible and Embedded Interaction* (pp. 159-162). https://doi.org/10.1145/1226969.1227003

Jackson, L. A., Witt, E. A., Games, A. I., Fitzgerald, H. E., Von Eye, A., & Zhao, Y. (2012). Information technology use and creativity: Findings from the Children and Technology Project. *Computers in human behavior*, *28*(2), 370-376.

Jayannavar, P., Narayan-Chen, A., & Hockenmaier, J. (2020). Learning to execute instructions in a Minecraft dialogue. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 2589-2602). https://doi.org/10.18653/v1/2020.acl-main.232

Johnson, M., Hofmann, K., Hutton, T., & Bignell, D. (2016). The Malmo platform for artificial intelligence experimentation. *Proceedings of the 25th International Joint Conferences on Artificial Intelligence* (pp. 4246-4247).

Lubart, T. I. (2001). Models of the creative process: Past, present and future. *Creativity Research Journal, 13*(3-4), 295-308.

Melián Díaz, D., Saorín, J. L., Carbonell-Carrera, C., & de la Torre Cantero, J. (2020). Minecraft: Three-dimensional construction workshop for improvement of creativity. *Technology, Pedagogy and Education, 29*(5), 665-678.

Niemeyer, D. J., & Gerber, H. R. (2015). Maker culture and Minecraft: Implications for the future of learning. *Educational Media International*, *52*(3), 216-226.

Kim, K. H. (2006). Can We trust creativity tests? A Review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, *18*(1), 3-14.

Kaufman, J. C., & Baer, J. (2012). Beyond new and appropriate: Who decides what is creative? *Creativity Research Journal, 24*(1), 83-91.

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A Three-level model of cognitive processing. *Human Development*, *4*, 222-232.

Lane, H. C., & Yi, S. (2017). Playing with virtual blocks: Minecraft as a learning environment for practice and research. In F. C. Blumberg, & P. J. Brooks (Eds.), *Cognitive Development in Digital Contexts* (pp. 145-156). Elsevier.

Lastowka, G. (2011). Minecraft as Web 2.0: Amateur creativity & digital games. In D. Hunter, R. Lobato, M. Richardson, & J. Thomas (Eds.), *Amateur media: Social, cultural and legal perspectives* (pp. 153-170). Routledge.

Mayer, R. E. (2019). Computer games in education. *Annual review of psychology*, *70*, 531-549.

Moreau, C. P., & Engeset, M. G. (2016). The Downstream consequences of problem-solving mindsets: How playing with LEGO influences creativity. *Journal of Marketing Research*, *53*(1), 18-30.

Peckham, M. (2013). *The 15 best Minecraft creations (and wildest destinations*). https://techland.time.com/2013/05/23/the-15-best-minecraft-creations-and-wildest-destinations/

Pusey, M., & Pusey, G. (2015). Using Minecraft in the science classroom. *International Journal of Innovation in Science and Mathematics Education*, 23(3), 22-34.

Sassenberg, K., Moskowitz, G. B., Fetterman, A., & Kessler, T. (2017). Priming creativity as a strategy to increase creative performance by facilitating the activation and use of remote associations. *Journal of Experimental Social Psychology, 68*, 128-138.

Sawyer, R. K. (2012). *Explaining creativity: The Science of human innovation*. Oxford University Press.

Schneider, B., Jermann, P., Zufferey, G., & Dillenbourg, P. (2010). Benefits of a tangible interface for collaborative learning and interaction. *IEEE Transactions on Learning Technologies*, *4*(3), 222-232.

Shabalina, O., Mozelius, P., Vorobkalov, P., Malliarakis, C., & Tomos, F. (2015). Creativity in digital pedagogy and game-based learning techniques; theoretical aspects, techniques and case studies. *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*. https://doi.org/10.1109/iisa.2015.7387963

Short, D. (2012). Teaching scientific concepts using a virtual world - Minecraft. *Teaching Science*, 58, 55-58.

Tichon, J. G., & Tornqvist, D. (2016). Video games: Developing resilience, competence, and mastery. In *Integrating Technology in Positive Psychology Practice* (pp. 247-265). IGI Global.

Watts, R. (2021). *The State of Minecraft in 2021.* PC Gamer. https://www.pcgamer.com/the-state-of-minecraft-in-2021/

Ward, T. B. (1994). Structured imagination: The Role of category structure in exemplar generation. *Cognitive psychology*, *27*(1), 1-40.

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954). A Factor-analytic study of creative-thinking abilities. *Psychometrika*, *19*(4), 297-311.

Examining the Effect of Socially Engaged Art Education with Virtual Reality on Creative Problem Solving

Hyojung Kim[1], Hyo-Jeong So[2*] and Ju-Yeon Park[3]
[1]Department of Fine Arts Education, Graduate School of Education, Ewha Womans University, Korea // [2]Department of Educational Technology, Ewha Womans University, Korea // [3]Cha Mirisa College of Liberal Arts, Duksung Women's University, Korea // hyojungk@ewha.ac.kr // hyojeongso@ewha.ac.kr // juyeonpark@duksung.ac.kr
[*] Corresponding author

ABSTRACT: The goal of this study was to examine the effect of engaging students in socially engaged art (SEA) education to create 3D virtual worlds for fostering creative problem-solving (CPS) skills. The study was conducted with 135 students (aged 16) of boys' high school in Korea who participated in the SEA program through four stages: Stage 1- appreciation and interpretation of artwork about social issues; Stage 2 - discussion on the potential solution to the selected social issue; Stage 3 - creating a 3D virtual world k to express proposed solutions; and Stage 4 - experiencing and sharing 3D virtual worlds. The following research questions guided the study: (1) What is the effect of SEA education with VR on students' CPS? (2) How are the students' CPS as expressed in their artifact (essay and VR work)? (3) What are the relationships between students' CPS and their artifact (essay and VR work)? For data collection, we administered the instrument to measure students' CPS skills in three areas (higher-order thinking, divergent thinking, and problem-solving) and also evaluated student essays and VR work to examine CPS specific to art education. Overall, the results indicate that the students improved their CPS skills significantly after participating in the SEA program. The CPS skills had significant relationships with the essay scores, whereas only one significant relationship was found between CPS and VR work. This study provides empirical findings concerning how the formal school curriculum can introduce students to an authentic context concerning social issues through artmaking practices with VR.

Keywords: Virtual reality (VR), Socially engaged art (SEA), Creative problem solving (CPS)

1. Introduction

Authentic learning environments present learners with complex real-life contexts and problems (Herrington & Oliver, 2000). Generally described as ill-structured or wicked problems, such authentic problems tend to be challenging to solve but stimulate students' creativity to go beyond well-known solutions. In K-12 education, pedagogical approaches that engage students in such authentic problem-solving activities (e.g., design thinking, problem-based learning, project-based learning, and service learning) have been emphasized in the school curricula to foster creative and critical thinking skills (Dorst, 2006; Hmelo-Silver, 2004).

In this study, we focus on a pedagogical approach called "socially engaged art" (SEA) in education as a particular mechanism for introducing authentic problems for learning through art practices and engaging students in a creative problem-solving process. SEA is viewed as a transpedagogy that blends "educational process and art making" (Helguera, 2011, p. 77). SEA has attracted considerable attention as a new direction for contemporary art education that expands the goal of art education beyond the traditional emphasis on teaching art skills and techniques disconnected from learners' lives. In SEA education, learners are engaged in meaning-making practices concerning various socio-cultural issues through participatory activities such as appreciation, critiques, and artwork creation.

With this backdrop, this study examines the effect of engaging students in SEA education to create 3D virtual worlds for fostering creative problem-solving (CPS) skills. In this study, we view VR as a relevant platform for students to express their creative ideas with the unique affordances of VR such as high representation fidelity and embodied actions (Dalgarno & Lee, 2010; Fowler, 2015). The existing literature on VR in the K-12 context has mainly focused on investigating how students consume VR content for cognitive learning (Maas & Hughes, 2020). The novelty of this study lies in that it examines the effect of VR as a tool for creation beyond consumption, positing students as a designer of VR content for expressing solutions to authentic social problems.

The following research questions guided this study: (1) What is the effect of SEA education with VR on students' CPS? (2) How are the students' CPS as expressed in their artifact (essay and VR work)? (3) What are the relationships between students' CPS and their artifact (essay and VR work)? By examining these questions,

this study aims to provide empirical findings concerning the effect of implementing the SEA approach with VR in the K-12 school context, which has been rarely reported in the existing literature.

2. Theoretical backgrounds

2.1. Creativity and creative problem solving

Creativity has been actively studied since the late 1950s. While there is no unified consensus on the concept of creativity, creativity has been approached as cognitive ability or affective disposition. Guilford (1967) sought to understand creativity through divergent thinking. He regarded creativity as the power to produce new and novel things, and not as a special talent of only certain people, but as an ability that everyone possesses. Torrance (1967) views creativity as the process of sensing gaps and formulating, testing, and retesting ideas to seek solutions. Sternberg and Lubart (1999) describe creativity as the confluence of multiple components such as intrinsic motivation, domain knowledge, and creativity-relevant skills (e.g., cognitive style, work style, and heuristics).

Scholars adopting cognitive approaches argue that creativity is manifested in problem-solving situations. Mayer (1989) defines creativity as the ability to solve problems that one has not previously learned to solve. Osborn (1953) built the foundation of CPS and proposed the seven-step CPS process: orientation, preparation, analysis, hypothesis, incubation, synthesis, and verification. Many variations of the CPS model do not deviate much from these steps. In CPS, learners are faced with complex problems and create solutions by identifying key factors and seeking new alternatives in a problem space. The difference between CPS and general problem solving (GPS) lies in whether iterative processes occur intentionally. Specifically, GPS is achieved by analyzing a problem space and then using existing knowledge to satisfy the solution requirements. While CPS goes through a similar process, it repeats the process of returning to the problem space to derive better solutions by reducing obstacles and constraints.

The iterative nature of CPS was further emphasized by Treffinger (1995) who developed the framework of CPS by moving away from fixed and sequential approaches. Treffinger suggested two important promises of CPS research. First, anyone might become creatively productive in meaningful ways and learn about their creative abilities. Second, CPS is not a simple, step-by-step process. With that, the CPS framework proposed by Treffinger includes three major components: understanding the challenge, generating ideas, and planning for action (Isaksen et al., 2000). Similarly, Mayer (1989) contends that teaching strategies to help students to be creative include (a) developing many component skills rather than a single monolithic general ability, (b) focusing on the process rather than the product of problem-solving, and (c) creative learning skills within specific content domains rather than as a separate course in general learning skills.

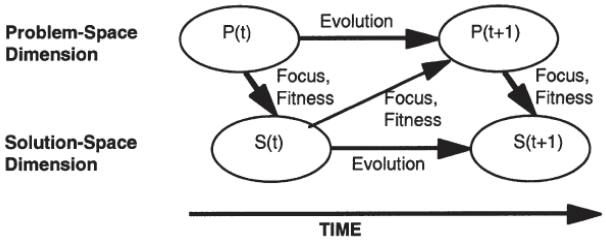
2.2. Creative problem solving in art education

By nature, art is a problem-solving process to create new artifacts, and creativity is the key driving force in artmaking. The co-evolution framework by Maher et al. (1996) helps understand how CPS processes unfold in art education. Traditionally, creativity has been regarded as a mysterious area in art and design education. Even those who are regarded as creative cannot identify significant events or factors that spurred their creative ideas due to the retrospective nature of such events (Wiltschnig et al., 2013). The co-evolution framework in Figure 1 suggests that creative design can be understood from two integrated dimensions: the problem-space and the solution-space. Maher et al. (1996) stated that the two spaces interact over time like evolution in the problem-design exploration process. The diagonal movement indicates that the problem leads to a solution (downward arrow), or the solution refocuses the problem (upward arrow). A fitness function indicates how close a given design solution is to achieving the current state. The definition of the problem can change according to the current state of the solution space, implying a co-evolution process.

Based on the co-evolution framework, Dorst and Cross (2001) studied nine industrial designers on their creativity through the think-aloud protocol. They found that designers used various approaches, such as analysis, synthesis, and evaluation to discover solutions, continuously crossing between the problem and solution spaces. Similarly, Maher and Tang (2003) studied the interaction between the problem and solution spaces through a protocol analysis of designers. They found that human designers had limited cognitive memory but strong reasoning between the problem requirements and the solution space. While these findings are interesting, the

application of the co-evolution framework has been limited to the study of professional designers. Research in the K-12 context, such as examining novice designers, is still lacking.

Figure 1. Co-evolution of the problem and solution spaces (Maher et al., 1996)



2.3. Socially engaged art education

As a pedagogical stance, SEA is built upon constructivist approaches to provide a new lens to examine the goal and power of art in education. Learners in SEA education are engaged in art practices "that are authentic to the ways artists in that field actually work and feel empowered enough in the situation to be willing and able to bring their own ideas to the process" (Wiggins, 2015, p.116). Helguera (2011) argued that traditional pedagogy in art education has failed to recognize three elements: (1) recognizing the creative performativity of the act of education, (2) the collective construction of an art milieu, and (3) knowing the artwork is not the end but is a tool for understanding the world. In essence, SEA in education emphasizes problem-solving through collaborative learning and the criticality of educational practices that are "collaborative and encourage cross-disciplinary dialogue and citizen engagement" (Rochielle & Carpenter, 2015, p. 131). Further, Schlemmer et al. (2017) succinctly pointed out the pedagogical role of SEA as stretching "beyond the production of aesthetically pleasing art objects to foster a dialog that integrates artistic practices, pedagogical processes, and creative possibilities in pursuit of a more equitable world" (p. 56). Hence, when SEA is integrated into the art education curricula, learning processes are often structured with collaborative, cross-disciplinary, and creative activities that engage students to create artwork under the themes of social change and civic engagement.

Despite the increased interest in SEA, research studies that examine the effect of SEA on student learning are still scarce. The existing literature is mostly qualitative and narrative. While a few studies are available, the existing studies on SEA education present an important message that learning in art education is not merely consuming content and developing art techniques but is creating value through artmaking practices. For instance, Roberts et al. (2008) reported the Storytelling Project curriculum where high school students were engaged in the critical examination of racism and social justice through storytelling and art. The students explored critical questions about racism expressed in the various forms of stories, such as historical documents and media with the theme of the "American dream," and then created counter-stories offering their imagination of new possibilities. The analysis of the student discourse revealed that art played a critical role in developing student agency and imagining alternatives. Chung and Li (2020) presented the possibility of integrating SEA for young learners. They attempted to teach elementary students about social justice issues in American society through artwork on the theme of homeless. The students were first exposed to the mural artwork by Skid Robert, who expressed the living conditions of homeless people, brainstormed ideas for alleviating the issues, and then finally drew a home for homeless people through printmaking. The study found that the students could discuss homeless issues meaningfully and critically as well as learning about art production skills.

2.4. Virtual reality for interpretation and expression

The preceding sections discussed the role of SEA in introducing authentic contexts and problems to learners and the mechanism of CPS from the co-evolution view. What is less elaborated in the literature on SEA and CPS is the space for learners to express their ideas and imagination, which we call an "interpretative and expressive space." In the design field, various technological tools and platforms support the process of CPS, especially dynamic interactions between the problem and solution spaces. For instance, Choi and Kim (2014) examined how the cognitive use of digital tools influenced the ability to derive creative design concepts among Korean university students majoring in design. Using digital tools for deriving ideas, the students expanded their creative thinking from a new viewpoint and transformed ideas using metaphors, analogies, and reasoning. Tark and Yoo (2018) found that VR as an expressive platform positively affected students' creative problem-solving ability and learning interest in social studies.

The present research is particularly interested in the affordances of VR as an interpretative and expressive space where students can express their creative ideas with 3D multimodal objects and images. The primary features of VR for learning are *immersion*, *real-time interaction*, and *reality*. The comprehensive literature review on the effect of immersive VR indicates that educational interventions with immersive VR produced significant advantages compared to non-immersive methods (Hamilton et al., 2021). Recent studies emphasize the importance of using VR to promote creative, comprehensive, and critical thinking in learners, beyond simple interest and novel experiences (Chang et al., 2020).

However, one promising area that has been less explored in the existing literature is to engage learners as the designer of VR content. While immersive experiences are beneficial for learners to explore virtual spaces, most VR-infused approaches tend to make learners passive users who experience the predesigned content. In the review of the literature on VR, AR, and MR in K-12 education, Maas and Hughes (2020) argue that most research used these advanced technologies for consuming materials and more research is needed to explore how students use these technologies as "a means of creation and discovery" (p. 245). Indeed, empowering students as a creator of VR content has pedagogical value in that students are engaged in complex problem solving through their creativity (Lim, 2008). Hu-Au and Lee (2017) contended that VR as a pedagogical tool presents several opportunities, such as providing authentic experiences, allowing new perspectives and empathy, and supporting creativity through visualization. When engaging learners to design VR content, it is important to present problems relevant to the students' interests and experiences to encourage them to actively participate with high motivation and play the role of creators (Choi et al., 2016). A recent study with junior high school students in Taiwan shows that design lessons with VR support the more engaging, exploratory, and reflective process of creative design than the lessons without the use of VR (Chang et al., 2020)

3. Methods

3.1. Research participants and context

This study was implemented at a boys' high school in a metropolitan city in Korea for about three months in 2019. The participants were 135 male students in the tenth grade (aged 16) who participated in six lessons in the SEA program for two months during the formal classroom hours of the art class. The designed SEA program included various activities to engage students to create artwork that visualizes their creative solutions to authentic social issues using VR as an interpretative and expressive space (see Section 3.2). The present research did not have a control group since the school policy was for the teacher to conduct all classes with the same lesson and activities. Prior to this study, the participants did not have any previous experiences with SEA programs and any VR authoring tools.

3.2. Lesson design and implementation

In this study, one high school teacher and two art education experts collaborated to design the SEA education program. The teacher had 10 years of teaching art in a high school. The two experts include one professor in art education with 20 years of teaching experience and another professor who teaches virtual reality in art education. In the SEA program design, a particular emphasis was placed on presenting authentic situations in which learners could realize the complexity of various social problems in society. An authentic context in this study refers to social problems in real-life situations and collaborative learning activities. Based on Maher's framework on the co-evolution of the design exploration, we designed learning activities to enable students to continuously navigate between the problem space and solution space to search for solutions. In the problem-space, the students were supposed to appreciate and unpack the messages that various artworks reflect, such as racism, environmental protection, and global warming. In the solution space, students used VR to create work that expressed the possible solutions to social problems.

Class activities were conducted face-to-face in a computer lab. Figure 2 presents the CPS activities in four stages. Stage 1 is the appreciation and interpretation of artwork about social issues. Stage 2 involves the discussion of the potential solution to the selected social issue. Stage 3 includes creating a 3D virtual world that expresses the proposed solution. In Stage 4, students experience and share the created 3D virtual worlds. Table 1 shows how the design of the SEA learning activities in each stage is guided by the design framework of authentic learning environments by Herrington and Oliver (2000).

In Stage 1, the students understood the relationship between art and society through various pieces of artwork. Table 2 presents the exemplary artwork used in the lesson to facilitate student discussion about authentic problems under three themes: political issues, natural environments, and social problems. Each artwork was presented digitally to the student through the projector in a computer lab. We used Feldman's (1992) art criticism model to guide critical thinking during the appreciation of various artwork. Feldman's model includes four steps in art criticism: description, analysis, interpretation, and judgment. Students did not need to perform the four-stage structure in a linear, sequential manner. Teachers engaged students to interpret artwork through their subjective experiences and to exchange ideas in class discussions.

Figure 2. Four stages of the socially engaged art (SEA) education program in this study

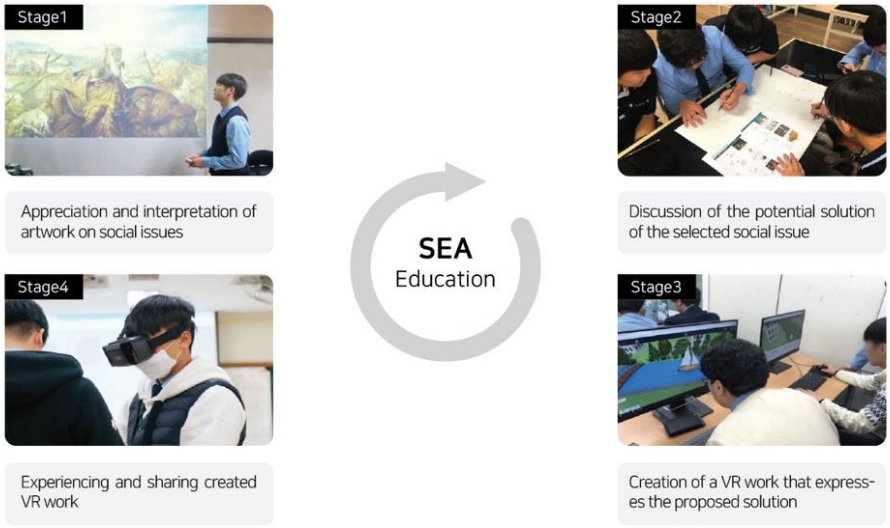


Table 1. Design guidelines of authentic learning environments implemented in this study

SEA education	Design guidelines (Herrington & Oliver, 2000)	Implemented in this study
Stage 1	• Provide authentic context and activities	• Students appreciate and interpret artwork on social issues to understand the relationship between art and society. • Students learn about Feldman's (1992) art criticism model as authentic activities by artists.
Stage 2	• Support collaborative construction of knowledge	• Students work in groups to brainstorm and discussion potential solutions to the selected social issue.
Stage 3	• Promote articulation and reflection • Provide coaching and scaffolding	• Students create a 3D virtual world to articulate and visualize their proposed solution. • Students write an essay about their VR work to reflect on their learning process. • The teacher provides necessary scaffolding to guide students to create a 3D virtual world in a computer-based environment
Stage 4	• Provide multiple roles and perspectives • Provide authentic assessment of learning within the tasks	• Students experience 3D virtual worlds created by other groups, switching their roles from a creator to a user. • Students are evaluated on the artifact created during the learning process rather than separate formal tests.

In Stage 2, the students formed groups of five to six members to select a particular social issue of their interest and discussed ideas for potential solutions face-to-face in a computer lab. Adopting the critical inquiry process by Geahigan (1999), the students were asked to search for external resources to organize and support their solutions. Brainstorming ideas on a large paper also facilitated the concretization of ideas and group discussions. The teacher carefully observed the process of group discussion and scaffolded students to solve problems by asking questions and providing necessary resources. For the iterative nature of CPS, the teacher helped the groups continuously navigate between the problem space and potential solutions and to reduce constraints in the problem space for deriving better solutions.

Stage 3 involves the creation of VR work to express the proposed solution. First, the students learned about the key functions of CoSpaces Edu, a web-based VR authoring software for easily creating 3D virtual worlds. With a

block-based programming language called CoBlocks, students new to coding or unfamiliar with programming could easily create a 3D virtual world. Students in groups built a virtual space by choosing 3D objects, such as cities, people, animals, and plants, and various materials and colors, adding animation effects through coding.

In Stage 4, the groups formally presented their 3D virtual world in the online space available in CoSpaces. This stage also allowed students to experience the 3D virtual worlds created by other groups using head-mounted displays (HMD). For the safety of students, they formed a pair, and one student was responsible for checking whether there was any danger in the physical space while the other student was experiencing the 3D virtual worlds with HMDs (model: LEAPmotion VR2).

Table 2. Themes of social problems expressed in the artwork

Themes	Theme 1: Artwork that reflects political issues (e.g., war, refugees, homelessness, racism & feminism)	Theme 2: Artwork that considers the natural environment (e.g., global warming & ecosystem)	Theme 3: Design that intends to solve social problems (e.g., universal design, nudge design, CPTED, green design & sustainable design)
Problem expressed in exemplary artwork			
	Racism	Global warming and deforestation	Environmentally friendly products

Note. CPTED: crime prevention through environmental design.

3.3. Data collection and analysis

3.3.1. Creative problem solving

We measured students' CPS skills through an instrument developed by Chi and Ju (2012). The instrument was developed to measure CPS skills as a general competency in all subject areas of the school curricula, reflecting the policy initiative by the Korean Ministry of Education to foster creatives leaders in a future society. We chose this instrument because it includes the core aspects of CPS emphasized in this program and was validated with 530 middle and high school students in Korea. The CPS instrument focuses on the cognitive dimension of creativity based on the theoretical perspectives of Guilford (1967) and Torrance (1967). Guilford (1967) associated divergent thinking with creativity and emphasized one's ability to generate multiple alternative solutions to a given problem. Torrance (1967) further elaborated that creativity is the process of sensing gaps and formulating, testing, and retesting ideas to seek solutions. Based on these theoretical perspectives, this instrument includes 13 items to measure the cognitive aspect of CPS on a 5-point Likert scale (1= Strongly disagree, 5=Strongly agree) in three areas: (a) higher-order thinking (4 items), (b) divergent thinking (5 items), and (c) problem-solving skill (4 items). Higher-order thinking is defined as an ability to logically analyze and synthesize what is learned. Divergent thinking measures an ability to deviate from a fixed frame for a problem and to derive several possible alternatives to generate novel and unique ideas. Problem-solving skill focuses on an ability to solve a given problem with diverse approaches and perspectives.

To ensure the validity of the instrument, we conducted reliability and factor analyses. The value of Cronbach's α was .897, which indicates a good internal consistency. The instrument was administered before and after the implementation of the designed SEA program. As mentioned earlier, since there was no control group in this study due to the school policy, we used a within-group comparison (pre-test and post-test) rather than a between-group comparison. A paired samples *t*-test was conducted to measure the effect of the SEA program on students' CPS.

3.3.2. Artifact evaluation: Essay and VR work

Since the above self-reported instrument could be subjective and measures CPS as a general competency, we collected additional qualitative data (i.e., essay and VR work) to examine students' CPS specific to the context of art education from multiple data sources. Developing a critical stance toward various social problems is one of the crucial goals advocated in SEA education. Hence, we collected and analyzed individual students' essays to measure the extent that students who participated in the designed SEA program developed their critical inquiry into art. The students were asked to write an essay about their VR work following the art criticism model by Feldman (1992): description, formal analysis, interpretation, and judgment. An essay template that guides the four elements of art criticism was provided to guide students' thinking processes.

We used the Art Criticism Assessment Rubric (ACAR) to evaluate student essays. Tam (2018) developed ACAR to evaluate art criticism by the Hong Kong Examinations and Assessment Authority. ACAR consists of eight elements for evaluating art criticism: (1) description, (2) formal analysis, (3) interpretation, (4) judgment, (5) application of aesthetic and contextual knowledge, (6) use of researched materials, (7) originality and balanced views, and (8) presentation, organization, and structure. The first four elements are based on Feldman's art criticism model, which is explained above. Next, the criteria "application of aesthetic and contextual knowledge" and "use of researched materials" in ACAR reflect Geahigan's inquiry-based art criticism model. Geahigan (1999) argues that art criticism is a critical inquiry process where students are engaged in searching for contextual knowledge and constructing their understanding of the artwork, beyond simple observation. Last two criteria, "originality and balanced views" and "presentation' measure students" intellectual writing skills. Since art appreciation and criticism theories by Feldman and Geahigan are included in the high school art curriculum in Korea and the students in this study were guided by these theories in the lessons, we decided that ACAR was suitable for this study as an evaluation rubric.

Some modifications to the original ACAR were made to make the rubric relevant to the context and purpose of the present study. The modified rubric has nine elements, as listed in Table 3. First, we used Elements 1 to 4 of ACAR that represent the four-stage structure of the art criticism model by Feldman (1992) to evaluate students' art criticism essays. This part mainly evaluates individual students' critical thinking processes and understanding of the problem-space dimension. Second, Elements 5 and 9 were used to evaluate students' VR work as the product of the solution-space dimension. Elements 5 to 7 were taken from Tam (2018) to mainly measure critical inquiry expressed in VR works. In addition, we created Element 8 (creative expression) and Element 9 (VR functions) to measure how the designed VR work represents students' creativity and the affordances of VR. The rubric uses a 10-point scale for each element: Very poor (2), Poor (4), Average (6), Good (8), and Excellent (10). Following the guideline by Tam (2018), the marks for (3) Interpretation and (5) Application of Aesthetic and Contextual Knowledge were doubled (total of 20 marks) as the core of higher-order thinking in art criticism.

Table 3. Elements and theoretical grounds of the evaluation rubric

Co-evolution in the design process	Elements	Evaluation target
Problem-space dimension	(1) Description (2) Formal analysis (3) Interpretation (4) Judgment	Art criticism essay (individual)
Solution-space dimension	(5) Application of aesthetic and contextual knowledge (6) Use of researched materials (7) Originality and balanced views (8) Creative expressions (9) Virtual reality functions	VR work (group)

Two raters (the teacher and one of the researchers) evaluated the student essays and VR work based on this rubric. Due to the large volume of data for evaluation, the teacher acted as the main rater to evaluate all data, whereas the researcher analyzed 30% of the data randomly selected from the pool. The inter-rater reliability was .980 for the student essays and .942 for the VR work, which indicates a high consistency between the two raters. Table 4 presents selected 3D virtual worlds based on the rubric scores in three levels (high, medium, and low) and QR codes to access each VR work.

Table 4. Examples of 3D virtual worlds by evaluation levels

	High	Medium	Low
Virtual reality work			
Problem	Water shortage in developing countries	Gender discrimination	Global warming
QR code to access VR work			

4. Results

4.1. Changes in creative problem-solving skills

First, factor analysis was conducted to verify whether the instrument accurately measures the three aspects of CPS. The suitability of the data for the factor analysis was determined by using the Kaiser Meyer Olkin (KMO) and Bartlett's test. The KMO values between .8 to 1.0 indicate that the sampling is adequate. As presented in Table 5, the KMO measure of sampling adequacy was .883, which is an acceptable value. The result of Bartlett's test of sphericity was statistically significant ($p < .05$). With that, we concluded that the data was suitable for the factor analysis. Next, exploratory factor analysis was performed with the principal axis method as the extraction criterion with varimax rotation. To determine the number of factors, we used the scree plot and Kaiser's criterion that the eigenvalue must be equal to or greater than 1.0. The cut-off value of factor loading was set at 0.4 or higher. After removing one item in higher-order thinking due to the low factor loading, we confirmed the three-factor structure of CPS (i.e., higher-order thinking, divergent thinking, and problem-solving skills).

Table 5. Kaiser–Meyer–Olkin (KMO) and Bartlett tests

Kaiser–Meyer–Olkin measure of sampling adequacy		.883
Bartlett's sphericity test	Approximate Chi-square	2351.321
	df	378
	Sig.	.000

Table 6. Descriptive statistics of pre-test and post-test for three variables of CPS

	Item	Pretest		Posttest	
		Mean	*SD*	Mean	*SD*
Higher-order thinking	(1) I fully understand what I have learned and apply it to other areas.	3.01	.96	3.53	.89
	(2) I logically analyze complex phenomena and grasp them as a whole.	2.76	.88	3.38	.91
	(3) I synthesize various pieces of information in context.	3.01	.88	3.58	.81
	Total	2.92	.75	3.50	.78
Divergent thinking	(4) I try new ideas or approaches to solve problems.	3.10	.92	3.61	.84
	(5) I do my assignments in a unique and individual way.	3.07	.87	3.63	.87
	(6) I tend to come up with a lot of ideas in a short time.	2.81	.97	3.51	.94
	(7) I tend to come up with a lot of new and original ideas.	2.95	.94	3.65	.84
	(8) I refine my thoughts and develop them into good ideas	3.08	.93	3.62	.88
	Total	3.00	.75	3.61	.68
Problem-	(9) I think and implement a solution to a problem	3.05	.84	3.67	.72

solving skills	from many angles.				
	(10) When faced with a difficult task to solve, I think of a number of alternatives.	3.30	.84	3.70	.82
	(11) I gather information related to a problem and draw a reasonable conclusion.	3.06	.86	3.74	.79
	(12) I expect the consequences of a solution to the problem in many ways.	3.08	.95	3.76	.79
	Total	3.12	.71	3.72	.64

Table 6 presents the descriptive statistics of students' CPS before and after participating in the SEA program. In the pre-test, the mean scores were 2.92 (*SD* = .75) for higher-order thinking, 3.00 (*SD* = .75) for divergent thinking, and 3.12 (*SD* = .71) for problem-solving skills. After participating in the SEA program, the students demonstrated increases in all three variables. In the post-test, the mean scores were 3.50 (*SD* = .78) for higher-order thinking, 3.61(*SD* = .68) for divergent thinking, and 3.72 (*SD* = .64) for problem-solving skills. Overall, the mean scores in each variable of CPS improved about 0.6 from the pre-test score.

Next, we conducted a paired samples *t*-test to examine the significance of the changes between the pretest and posttest scores. As listed in Table 7, the differences were statistically significant for all three variables: higher-order thinking ($t = 6.992$, $p < .05$), divergent thinking ($t = 9.324$, $p < .05$), and problem-solving skills ($t = 7.908$, $p < .05$).

Table 7. Paired samples *t*-test results for each of the three variables in CPS

Posttest-pretest	Mean	*SD*	*SE*	*t*	*df*	*p*-value
Higher-order thinking	.57	.94	.08	6.992*	134	.00
Divergent thinking	.60	.75	.06	9.324*	134	.00
Problem solving skills	.60	.88	.07	7.908*	134	.00

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

4.2. Artifact evaluation

Table 8 presents the scores of the artifact evaluation. First, in the essay evaluation based on the rubric, the mean values were 6.66 (*SD* = 2.13) in description, 6.74 (*SD* = 2.23) in formal analysis, 13.67 (*SD* = 4.65) in interpretation, and 6.16 (*SD* = 2.43) in judgment. Overall, the results revealed that student scores in the four art criticism areas were slightly above the mid-point of the scale. The mean for the total score was 33.23 (*SD* = 10.39) out of 50. Second, the evaluation of the VR work indicates that the score for creative expression was the highest among all the elements whereas the score for the use of researched materials was the lowest. Specifically, the mean values were 13.84 (*SD* = 3.79) for the application of aesthetic and contextual knowledge, 6.22 (*SD* = 1.96) for the use of researched materials, 6.77 (*SD* = 1.85) for originality and balanced views, 7.02 (*SD* = 2.48) for creative expressions, and 6.53 (*SD* = 2.05) for VR functions. The mean for the total score was 40.39 (*SD* = 10.92) out of 60.

Table 8. Descriptive statistics for the rubric evaluation of essay and VR work ($n = 135$)

	Element (total mark)	Min	Max	Mean	*SD*
Art criticism essay	Description (10)	2.0	10.0	6.66	2.13
	Formal analysis (10)	2.0	10.0	6.74	2.23
	Interpretation (20)	4.0	20.0	13.67	4.65
	Judgment (10)	2.0	10.0	6.16	2.43
	Total (50)	10.0	50.0	33.23	10.39
VR work	Application of aesthetic and contextual knowledge (20)	4.0	20.0	13.84	3.79
	Use of researched materials (10)	2.0	10.0	6.22	1.96
	Originality and balanced views (10)	2.0	10.0	6.77	1.85
	Creative expressions (10)	2.0	10.0	7.02	2.48
	Virtual reality functions (10)	2.0	10.0	6.53	2.05
	Total (60)	16.0	58.0	40.39	10.92

4.3. Correlations

As shown in Table 9, we analyzed the correlations between the students' posttest scores on CPS and the scores of the artifact produced during the learning process (i.e., essay and VR work scores). The reason for conducting the correlation analysis was to examine how CPS perceived by individual students is related to the output produced at the individual level (essay) and the group level (VR work). Regarding the relationship between CPS and essay scores, statistically significant correlations exist among all variables, except the relationship between higher-order thinking and judgment. The highest correlation was found in the relationship between divergent thinking and the description (r = .213, p < .05). We also conducted a correlation analysis between the students' posttest scores on CPS and their scores on the VR work. Only one statistically significant correlation was found among these variables, which was different from the trend observed in the correlation with the essay scores. The only significant correlation was between creative expressions and divergent thinking (r = .175, p < .05).

Table 9. Correlations between CPS, essay, and VR work scores

		Creative problem solving (CPS)		
		Higher-order thinking	Divergent thinking	Problem-solving skills
Art criticism essay	Description	.185*	.213*	.184*
	Formal analysis	.186*	.193*	.171*
	Interpretation	.203*	.170*	.205*
	Judgment	.160	.185*	.210*
VR work	Application of aesthetic and contextual knowledge	-.004	.089	-.006
	Use of researched materials	.019	.079	-.022
	Originality and balanced views	-.053	.101	-.028
	Creative expressions	.011	.175*	.021
	Virtual reality functions	-.003	.119	-.019

Note. *p < .05.

5. Discussion and conclusion

5.1. Discussion of key findings

In this study, we examined the effect of engaging students in SEA education to introduce authentic contexts for learning and fostering CPS skills beyond well-known solutions. In particular, we engaged 135 high school students in Korea in creating 3D virtual worlds as an interpretive and expressive space to represent their solutions in an immersive VR platform. This section revisits the three research questions that guided the present study and discusses the implications.

Regarding the first research question, the changes in CPS scores measured in the pretest and posttest were statistically significant, indicating that the students improved their ability to creatively solve problems significantly after participating in the SEA program. This finding is consistent with the existing SEA research in art education that has reported positive effects on students' learning outcomes (e.g., Chung & Li, 2020; Roberts et al., 2008). We attribute this positive effect to the intentional design of the SEA activities, which aims to promote students' CPS skills in four inter-related stages. The learning activities in each stage were designed following the framework of authentic learning environments (Herrington & Oliver, 2000). Such intentional design also guided the students to navigate between the problem space and the solution space during the problem-solving process. In particular, the SEA program highly emphasized building empathy concerning various social problems by appreciating and discussing the meaning of artworks. One of the core implications of authentic learning is that when learning is decontextualized from students' daily life, students have difficulty building empathy concerning social issues and tend to believe that knowledge is distant from their lives (Anderson et al., 1996). The first stage in the SEA program provided students with an opportunity to see the relevance of various social issues to their personal lives. Further, the group discussion and class debate provided a platform where the students unpacked the meaning of the artwork from more critical stances. With a sufficient discussion regarding how serious and important each social problem is, students could extend their thinking with relevance and empathy, which are important attributes of designer ways of knowing (Cross, 2007).

The second research question examined how CPS skills are expressed in the concrete artifacts (i.e., essay and VR work) that students created. Adopting the co-evolution framework by Maher et al. (1996), we used the comprehensive rubric to examine students' CPS in the problem-space dimension captured as process narratives

in writing and their CPS in the solution space captured as a visualized expression in VR work. The evaluation indicates that the mean of the art criticism essays was 33.23 (*SD* = 10.39) out of 50 whereas the mean of the VR work was 40.37(*SD* = 10.92) out of 60. The scores were not as high as we expected. The standard deviations were also rather high, indicating that the levels of the student artifacts were diverse. We speculate that the essay scores were influenced by the individual students' critical and analytical writing abilities. While the essay template included statements about what each element of the art criticism model requires, art criticism is a challenging activity even for university students majoring in art (Wolff & Geahigan, 1997). The finding implies that students may need more scaffolding to perform critical and analytical writing, especially for those who received low scores on the essay. Concerning the evaluation of the VR work, the results show a rather high standard deviation indicating group differences. Given that this was the first implementation of the SEA program with VR, the finding suggest that the students may need more exposure and experience to transfer CPS skills to a virtual platform.

In the last research question, we intended to examine whether any statistically significant relationships exist between the CPS skills and the artifact produced during the learning process (i.e., essay and VR work scores). The overall results indicate that the CPS skills had significant relationships with the essay scores except for one relationship, whereas only one significant relationship exists between the CPS skills and VR work. Our finding is rather different from the previous study such as Chang et al. (2020) and Hu-Au and Lee (2017) that found positive effects of VR on the creative design process and outcomes. We interpreted the different results from the perspective of individual and group creativity. While both the CPS skills and essays were measured at the individual level, the VR work was measured at the group level. The literature suggests that individual creativity and group creativity should be understood as different entities (Sawyer, 2010). That is, group creativity cannot be reducible to individual-level explanations. Likewise, the paradigm of knowledge creation has suggested that a group is more than the sum of individuals, emphasizing the power of collective intelligence (Bereiter & Scardamalia, 2014; van Aalst, 2009). Individual CPS skills may be limited in the VR production due to group dynamics. However, the significant relationship between creative expressions in the VR work and the individual students' divergent thinking implies some association between individual creativity and group creativity. This finding suggests the need to scaffold the CPS process at a group level so that individual students in a group could express their creative ideas freely to create synergy for the final solutions. About the affordances of VR, it is encouraging to see the significant relationship between creative expressions and divergent thinking. This may imply the importance of the 3D VR platform as a space to express creative ideas in divergent ways, which is rather limited in paper-based or 2D platforms.

5.2. Limitations and areas for future research

Some limitations of this study include the following. First, this study measured individual creativity but not group-level creativity. Because the VR production was done in groups, it would be useful for future research to examine how group creativity unfolds in the CPS process. Future research can employ discourse analysis, which has been used in the existing literature on designers' discussions (e.g., Dorst & Cross, 2001), to unpack the nature of group creativity. Second, the generalization of the findings should be limited to a similar research context and student profiles. Since the study was conducted at a boys' high school, future research needs to examine whether similar findings can be obtained with female students. In addition, future implementations need to consider students' technology competency for VR content production. While CoSpaces Edu is a user-friendly program for creating 3D virtual worlds, learners with little or no programming skills may need additional technical training. Third, the present study did not compare the effect of the SEA with VR to other approaches due to the school policy. We suggest that future research needs to conduct an experimental study with a control group that uses a traditional approach without the support of advanced technologies like VR. Lastly, this research focused on the cognitive aspect of creativity within problem-solving situations and did not consider affective dimensions of creativity. One of the promising areas for future research is to investigate the interplay of cognitive and affective dimensions of CPS with the consideration of students' affective attributes such as curiosity, openness, sensitivity, and persistence.

Although SEA has received much attention as a new direction for art education, scarce empirical research has examined how SEA programs affect students' creative learning. The present study supports that engaging students in solving authentic social problems through VR creation is a promising approach to facilitating students' CPS skills. The study findings also provide insight into the importance of engaging students in creating social values through VR beyond simply consuming VR content. We hope this study can provoke more research interest in the influence of socially engaged practices in other disciplines.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the Korea Foundation for the Advancement of Science and Creativity in 2019.

References

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11.

Bereiter, C., & Scardamalia, M. (2014). Knowledge building and knowledge creation: One concept two hills to climb. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge creation in education* (pp. 35-52). Springer.

Chang, Y. S., Chou, C. H., Chuang, M. J., Li, W. H., & Tsai, I. F. (2020). Effects of virtual reality on creative design performance and creative experiential learning. *Interactive Learning Environments*, 1-16. https://doi.org/10.1080/10494820.2020.1821717

Chi, E., & Ju, U. (2012). Exploring the construct and developing the scale for the measurement of creative leader competency. *Journal of Educational Evaluation, 25*(1), 69-94.

Choi, D. H., Dailey-Hebert, A., & Estes, J. S. (Eds.). (2016). *Emerging tools and applications of virtual reality in education*. Information Science Reference.

Choi, H. H., & Kim, M. J. (2014). Digital utilization for creative problem-solving in design education with a focus on cognitive aspect of the digital. *Journal of the Korean Society Design Culture, 22*(3), 599-608.

Chung, S. K., & Li, D. (2020). Socially engaged art education: Exploring issues of homelessness in an elementary art classroom. *International Journal of Education & the Arts, 21*(21). http://doi.org/10.26209/ijea21n21.

Cross, N. (2007). From a design science to a design discipline: Understanding designerly ways of knowing and thinking. In R. Michel (Ed.), *Design research now* (pp. 41-54). Birkhäuser.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10-32.

Dorst, K. (2006). Design problems and design paradoxes. *Design Issues, 22*(3), 4-17.

Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. *Design Studies, 22*(5), 425-437.

Feldman, E. B. (1992). *Varieties of visual experience*. Prentice-Hall.

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

Geahigan, G. (1999). Models of critical discourse and classroom instruction: A Critical examination. *Studies in Art Education, 41*(1), 6-21.

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8*(1), 1-32.

Helguera, P. (2011). *Socially engaged art.* Jorge Pinto Books.

Herrington, J., & Oliver, R. (2000). An Instructional design framework for authentic learning environments. *Educational Technology Research and Development, 48*(3), 23-48.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266.

Hu-Au, E., & Lee, J. J. (2017). Virtual reality in education: A Tool for learning in the experience age. *International Journal of Innovation in Education, 4*(4), 215-226.

Isaksen, S. G., Dorval, K. B., & Treffinger, D. J. (2000). *Creative approaches to problem solving: A Framework for change*. Kendall Hunt Publishing Company.

Lim, C. P. (2008). Spirit of the game: Empowering students as designers in schools? *British Journal of Educational Technology, 39*(6), 996-1003.

Maas, M. J., & Hughes, J. M. (2020). Virtual, augmented and mixed reality in K–12 education: A Review of the literature. *Technology, Pedagogy and Education, 29*(2), 231-249.

Maher, M. L., Poon, J., & Boulanger, S. (1996). Formalising design exploration as co-evolution. In J. S. Gero & F. Sudweeks (Eds.), *Advances in formal design methods for CAD: Proceedings of the IFIP WG5.2 Workshop on Formal Design Methods for Computer-Aided Design* (pp. 3-30). Springer.

Maher, M., & Tang, H. H. (2003). Co-evolution as a computational and cognitive model of design. *Research in Engineering Design, 14*(1), 47-64.

Mayer, R. E. (1989). Cognitive views of creativity: Creative teaching for creative learning. *Contemporary Educational Psychology, 14*(3), 203-211.

Osborn, A. F. (1953). *Applied imagination: Principles and procedures of creative thinking*. Scribners.

Roberts, R. A., Bell, L. A., & Murphy, B. (2008). Flipping the script: Analyzing youth talk about race and racism. *Anthropology & Education Quarterly, 39*(3), 334-354.

Rochielle, J., & Carpenter, B. S. (2015). Navigating the third space. *Journal of Curriculum and Pedagogy, 12*(2), 131-133.

Sawyer, R. K. (2010). Individual and group creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 366-380). Cambridge University Press.

Schlemmer, R. H., Carpenter, B. S., & Hitchcock, E. (2017). Socially engaged art education: Practices, processes, and possibilities. *Art Education, 70*(4), 56-59.

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

Tam, C. O. (2018). Evaluating students' performance in responding to art: The Development and validation of an art criticism assessment rubric. *International Journal of Art & Design Education, 37*(3), 519-529.

Tark, J. S., & Yoo, M. H. (2018). The Effects of creativity convergence program utilizing virtual reality creation platform on elementary school students' creative problem solving ability, 21st century skills and learning interest about social subject. *The Journal of Korean Practical Arts Education, 24*(4), 73-101.

Torrance, E. P. (1967). Epilogue: Creativity in American education. In J. C. Gowan, G. D. Demos, & E. P. Torrance (Eds). *Creativity: Its educational implication* (pp. 319-332). John Wiley.

Treffinger, D. J. (1995). Creative problem solving: Overview and educational implications. *Educational Psychology Review, 7*(3), 301-312.

van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4(*3), 259-287.

Wiggins, J. (2015). Constructivism, policy, and arts education. *Art Education Policy Review, 116*(3), 115-117.

Wiltschnig, S., Christensen, B. T., & Ball, L. J. (2013). Collaborative problem–solution co-evolution in creative design. *Design Studies, 34*(5), 515-542.

Wolff, T. F., & Geahigan, G. (1997). *Art criticism and education*. University of Illinois Press.

Using the Online Self-Directed Learning Environment to Promote Creativity Performance for University Students

Xinquan Jin[1], Qiang Jiang[1*], Weiyan Xiong[2], Xingzhu Pan[1] and Wei Zhao[1]
[1]School of Information Science and Technology, Northeast Normal University, China // [2]Institute of Policy Studies, School of Graduate Studies, Lingnan University, Hong Kong // jinxq805@nenu.edu.cn // jiangqiang@nenu.edu.cn // weiyanxiong@ln.edu.hk // panxz537@nenu.edu.cn // zhaow577@nenu.edu.cn
[*]Corresponding author

ABSTRACT: Creativity has been identified as a critical educational goal and an essential 21st-century skill, which can be captured through learning capabilities, thinking skills, and academic achievement. Although the relationship between creativity performance and self-directed learning (SDL) was theoretically researched, few studies have thoroughly investigated the exact nature of this association from a practical perspective. Therefore, this study aimed to design an online self-directed learning environment (OSDLE) to improve students' creativity performance. The OSDLE was proposed with functions such as planning, learning, evaluation, and reflection, based on the three dimensions of personal attributes, process, and learning context. A quasi-experimental study was conducted in a university in Northeast China to explore the influence of the OSDLE on creativity performance. One hundred and six university students as study participants were randomly assigned to experimental and control groups. Participants in the experimental group learned in the OSDLE, whereas those in the control group learned in traditional classroom methods. The results indicated that the students using the OSDLE exhibited significant improvements in creativity performance. Furthermore, the SDL capabilities of the experimental group demonstrated gradual and continuous improvement. In addition, students' thinking skills and academic achievement in the experimental group were higher than those of the control group. The main findings together are discussed in depth.

Keywords: Creative learning, Self-directed learning, Creativity performance, Online self-directed learning environment

1. Introduction

Creativity is considered as one of the crucial competencies required for students to survive and thrive in the 21st century (Hong & Song, 2020; Hernández-Torrano, & Ibrayeva, 2020). The rapid advancement of technologies has increased active learning opportunities for novices and experts, and the complexity of this landscape means that creative students must become independent learners and exhibit a general trait of self-direction (Tekkol & Demirel, 2018; Garrison, 1997). When students are increasingly expected to be responsible for their learning, self-directed learning (SDL) can serve as an essential component of creative activity (Lemmetty & Collin, 2021; Lee, 2019; Yeh & Lin, 2015). Although it has been repeatedly asserted that creative experiences and achievements are associated with SDL (Morris, 2020; Torrance & Mourad, 1978), the exact nature of this association remains unclear. Therefore, examining the possibility that creative learning outcomes can be supported through SDL is an urgent concern (Gralewski & Karwowski, 2019).

Amabile (1982) stated that the outcome of the creative process and the process of creativity could be applied to facilitate the measurement of creativity interchangeably. From the outcome perspective, thinking skills are mainly identified as originality, flexibility, and fluency of thinking, generally considered reliable indicators of creativity (Guilford, 1967). Empirical studies have proved that thinking skills and the creation of creative products are positively correlated (Hardy et al., 2017; Morris, 2020). Hence, improving students' thinking skills is essential to support the individuals' development in creativity performance.

Meanwhile, a widespread belief is that creativity relies on a learner's knowledge, which views as an information source for creativity (Amabile, 1982). Investigations have indicated a positive correlation between creativity performance and academic achievement. Relevant knowledge is the foundation for scientific creative activities, including identifying scientific problems, designing scientific experiments, and more (Klahr & Dunbar, 1988). Hence, students with expertise in a field can better retrieve information they need and make connections to other information previously learned, which lays a solid foundation for creativity.

Notably, SDL capabilities can be viewed as a prerequisite for promoting creativity performance (Lemmetty & Collin, 2021; Morris, 2020). Lee (2019) used repeated measures mixed model analysis and hierarchical linear model analysis to demonstrate that SDL improved students' creative abilities, but the explanation of this

relationship is unclear. In this regard, Morris (2020) conducted a literature review and highlighted that SDL supports creative learning outcomes. Specifically, when individuals with SDL capabilities can be open to new ideas, they may be able to view knowledge in a different and creative way (Toh & Kirschner, 2020).

In addition, Smith (2009) emphasized that online platforms can facilitate an environment that supports independent learning and enhances participative behaviors and possibly even creativity. According to Mishra et al. (2013), students in school lack the opportunity to navigate complex exploration and creativity performance because of the bounded designs of the traditional classroom. They theoretically stated that "open-ended, technology-rich learning contexts appear to provide opportunities for students to be structured in their ways of thinking" (Mishra et al., 2013), which is the crux of creativity. Although the online environment can help students enrich their learning experience, an unwanted consequence is that it can disturb students' creativity processes if they lack proper SDL capabilities. Based on the studies of the role of SDL practices in creative activity in a technological context (e.g., Lemmetty & Collin, 2021), online learning environments are proved to be more effective at promoting SDL (Candy, 1991), which can provide opportunities for students to foster innovative thinking (Mishra et al., 2013). However, it is no longer enough to simply explore the improvement of creativity by the SDL environment from a theoretical perspective. Therefore, it is necessary to provide an online self-directed learning environment (OSDLE) in school settings to prepare students for creativity during the learning process.

In this study, an OSDLE was designed with four modules, and its effects on students' creativity performance were assessed through analyzing their SDL capabilities, thinking skills, and academic achievement. Section 2 reviews the literature on creativity in online learning and presents the conceptual framework for SDL in the online learning environment to situate the study. Section 3 describes the OSDLE developed in this study. In Sections 4 and 5, the experiment and results evaluation are demonstrated. Finally in Section 6, after the research findings are discussed, conclusions and suggestions for future research are presented.

2. Literature review

2.1. Creativity

Creativity plays a crucial role in further developing human intellectual potential, and the interest in creativity within the scope of education has grown exponentially (Hernández-Torrano, & Ibrayeva, 2020). The growth in interest is due mainly to creative thinkers being able to adjust easily to new situations and create new and original ideas that are considered useful or valuable by integrating the knowledge, skills, and experience of diverse professional fields (Chiu & Tu, 2014; Rhodes, 1961). Therefore, the concept of creativity is complex and challenging to define in the research field because creativity is no longer a single attribute, but rather a set of attributes (Sternberg, 2006). According to Ma (2009), personal factors can be applied to define a person's creativity, including personality, cognitive ability, thinking style, and academic achievement.

Previous studies have analyzed person-centered variables that contribute to creativity. From a more holistic perspective, many studies on creativity have adopted approaches to investigate different aspects of creativity, such as Rhodes's (1961) "four P's of creativity," which means person, process, product, and press. Isaksen et al. (1993) explained the simultaneous interaction among "four P's" components. This model highlights the "creative process" as an integral part, emphasizing the interactions among the components related to individuals and environments. Therefore, the generation or manifestation of creativity performance can be captured through SDL capabilities, thinking skills, and academic achievement.

Since its inception, creativity research has been linked to education (Hernández-Torrano, & Ibrayeva, 2020). Fostering creativity has attracted much attention in the field of education (Chiu & Tu, 2014). Creativity can be regarded as a practice-based process wherein it is essential for students to actively engage in the creative process (Dewey, 1916). Consequently, the practices of SDL appear to match these demands. Tekkol and Demirel (2018) used the survey method and concluded a moderate positive relationship between SDL and creativity. This relationship is also found in Lee (2019) and Lemmetty and Collin (2021). However, few attempts have been made to design the OSDLE to improve creativity. To cultivate students' creativity, constructing an environment that enables students to SDL and implement learning strategies flexibly to generate new ideas is necessary. Therefore, the OSDLE can be considered an effective way to promote creativity performance.

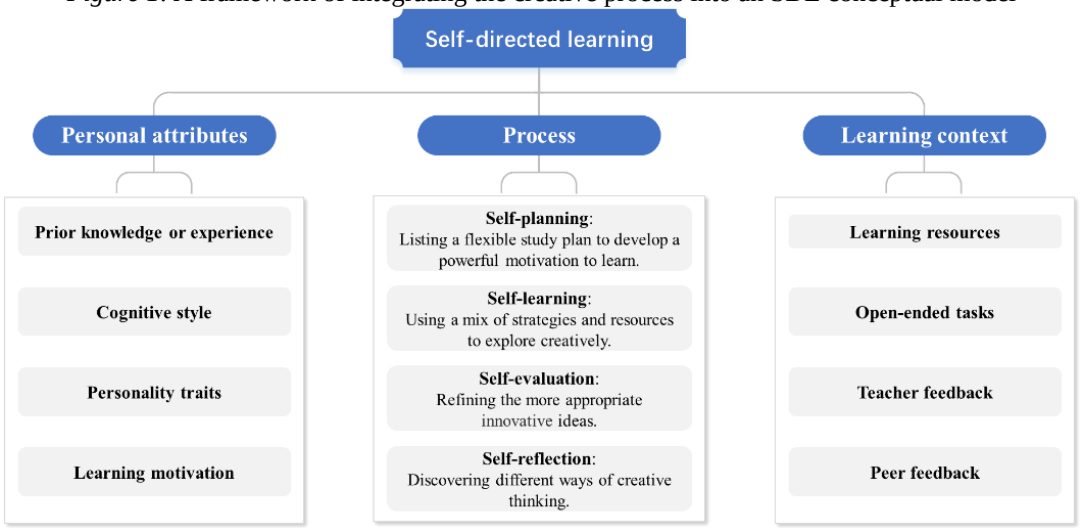
2.2. Self-directed learning

SDL was previously defined as the process of an individual actively learning with or without the assistance of others (Knowles, 1975). Guglielmino (1977) proposed another definition and posited that personal attributes determine whether an individual has the ability and potential of SDL. The definition of SDL has also been accompanied by the concept of self-regulated learning, yet SDL is a broader concept that involves the use of self-regulated learning strategies, including planning, monitoring, and evaluation (Dickinson, 1987; Manganello et al., 2019; Rubenstein et al., 2018). However, given that a student does not learn nor act individually, previous studies on SDL showed that the external learning context could play a role in SDL development (Chu et al., 2012; Kim et al., 2021; Mamun et al., 2020). Therefore, SDL commonly considers not only the process and personal attributes but also the importance of learning context.

Subsequently, numerous conceptual models for SDL were developed to better understand and foster SDL in the learning environment, including Candy's (1991) Four-Dimensional Model, Garrison's (1997) Three-Dimensional Model, and Song and Hill's (2007) Conceptual Model. Despite the differences among these models, they are significantly overlapped regarding the critical constructs associated with each model (Morris & Rohs, 2021). In most of the SDL models reviewed, personal attributes, process, and learning context were discussed to a certain extent (Song & Hill, 2007).

Regarding creativity, Song and Hill's (2007) model places greater emphasis on the online learning context factor, and the clear learning process, which may more accurately introduce a conceptual model to understand SDL in an online environment while creating. Moreover, recent research indicated that students need to have a high level of SDL capabilities to successfully develop the ability to think creatively (Morris, 2020). SDL enables individuals to change their mode of learning from "passive study" to "independent study," thereby improving their creativity performance. Therefore, we present the conceptual framework of SDL and describe how one can facilitate creativity during the process (Figure 1).

Figure 1. A framework of integrating the creative process into an SDL conceptual model



2.2.1. Personal attributes

Personal attributes are described as characteristics of students in a specific learning situation (e.g., prior knowledge or experience, cognitive style, personality traits, and learning motivation). Specifically, the relationship between personal attributes and creativity has been assessed by many scholars (Amabile, 1982; Sternberg, 2006). Suppose the level of students' personal attributes is high. In that case, they tend to retrieve knowledge better, exhibit independent judgment, are highly self-disciplined, and remain enthusiastic for learning (Morris & Rohs, 2021), making it easier for students to perform creative behaviours or achieve innovative results outcomes.

2.2.2. Process

The process focuses on students' autonomous learning process (Song & Hill, 2007), including self-planning, self-learning, self-evaluation, and self-reflection stages. During the self-planning phase, students list a flexible study plan to identify learning goals and a way forward. It stimulates each student to develop a powerful motivation to learn (Tang et al., 2020). The self-learning phase provides an opportunity for students to creatively explore the task at their pace using a mix of strategies and resources (Hardy et al., 2017). During testing and monitoring, students move through phases of self-evaluation and self-reflection. These phases refine the more appropriate innovative ideas and require students to critically evaluate their decisions when reflecting on the learning process (Yeh & Lin, 2015), which can support students in discovering different ways of creative thinking.

2.2.3. Learning context

The learning context has various factors that can affect the development of SDL capabilities, including the learning resources, open-ended tasks, and feedback from the teacher and peers (Song & Hill, 2007). Sufficient learning resources and open-ended tasks permit students to access updated cognitive. Additionally, students may benefit from critical and constructive feedback, which in turn contributes to promoting creativity performance.

2.3. Research motivation and questions

In light of the literature reviewed, despite a convincing theoretical rationale that creativity and SDL are positive attributes (Mishra et al., 2013; Lemmetty & Collin, 2021), research on teaching practicum-relevant outcomes that can be included in SDL to promote creativity is scarce. Moreover, whereas prior researches have indicated that the nature of creativity is definitionally difficult to capture and identify (Gralewski & Karwowski, 2019; Ness, 2012), less attention has been given to understanding creativity-related variables, such as learning capabilities, thinking skills, and academic achievement. Therefore, the study expands on previous research focusing on improving creativity as a motivational consequence of designing an OSDLE. The assessment of creativity demands multiple avenues of measurement because it is a multidimensional concept. Therefore, this study investigates students' creativity by exploring and examining their SDL capabilities, thinking skills, and academic achievement. The present study was guided by the following research questions:

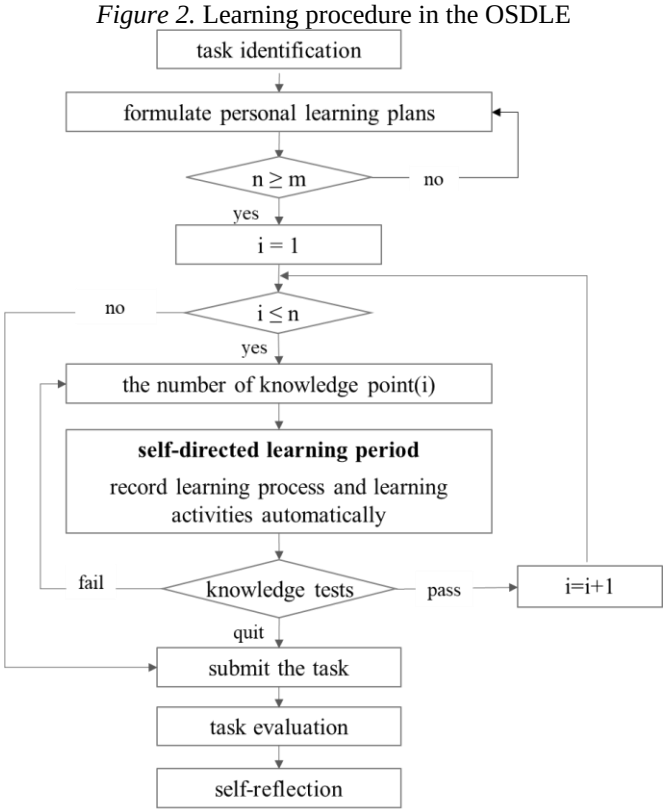
- RQ1 - Do students enhance their SDL capabilities by learning in the OSDLE?
- RQ2 - Do students who learn in the OSDLE show better thinking skills than those who learn in traditional classroom methods?
- RQ3 - Do students who learn in the OSDLE show better academic achievement than those who learn in traditional classroom methods?

3. Design of online self-directed learning environment

3.1. Learning procedure in the OSDLE

The learning procedure of the proposed OSDLE is shown in Figure 2, where *m* is the minimum number of knowledge-learning points required to be learned set by the course teacher to complete the specific learning task. Students first view the learning task and formulate a definite learning plan based on their learning experience and previously accumulated knowledge, where *n* is the number of planned knowledge-learning points set by students. Then, each knowledge learning content is embedded in the Q&A module and an evaluation module. The tests consist of a set of two-tier multiple-choice questions. Each test item has three or four choices in the first tier, and there are three or four reasons for each choice in the second tier (Yang et al., 2015). Each test item was developed and reviewed by domain experts and researchers (r = 0.88). To consider a question as correctly answered, students need to be answered correctly in both tiers. An example of a two-tier test item is shown in Appendix Figure 11 and Table 5. After passing the knowledge test, students can enter the next knowledge point for learning; otherwise, a continued revision of this knowledge point is suggested until the test is passed. Moreover, if students want to quit further learning, they can choose to submit the task directly. Additionally, students could evaluate peers' tasks, observe excellent artifacts, and share reflection logs based on the learning behaviour assessment form.

Figure 2. Learning procedure in the OSDLE



3.2. OSDLE function module

The OSDLE includes four modules: planning, learning, evaluation, and reflection. Students make their learning plans based on learning context and tasks. Then, the OSDLE receives students' learning plans from the planning module, deploys the tasks on the learning module, monitors their progress, reconfigures the learning tasks based on learning needs and evaluation results, and reports status and results to the evaluation module. The following subsections demonstrate the details of the modules in the OSDLE.

3.2.1. Planning

In the planning module, students need to fill in the K-W-L (know, want to, learn) chart, where K means what we know, W means what we want to know, and L means what we learned and still need to learn (Ogle, 1986). During step K, the teacher can discover what the students do not know, and then provide relevant learning materials. Step W helps students develop clear personal goals. The majority of step L involves promoting students' reflection. To lay the foundation for calculating SDL capabilities, students also need to set up the planned learning time and knowledge-learning points. In general, when students decide on the need for guidance, they may overcome procrastination in online learning (Shadiev et al., 2018), which urges students to create new ideas actively.

3.2.2. Learning

The learning module consists of three sub-modules: learning content, monitoring, and Q&A. To support students in identifying their academic strengths and weaknesses, the OSDLE divides the learning content module into three parts: learning, learned, and not learned.

Extensive knowledge or ability in learning is the basic foundation for creativity performance (Amabile, 1982). When using the OSDLE, students are not passive recipients of knowledge, but rather seek the proper learning resources according to their needs, enabling them to have the opportunity to acquire more diversified knowledge, which may lead to creative outcomes.

Students' cognitive structure in a specific field will be developed and improved further when actively learning task-related knowledge. Drawing from Bloom's (1956) Taxonomy of Educational Objectives, the learning content consists of three levels: knowledge, comprehension, and application. Based on the Classical Testing Theory (Holland & Hoskens, 2003), each student's level of learned content is ranked on a two-level ordinal scale: complete content with a green flag and incomplete content with a red flag, as shown in Figure 3.

The monitoring module displays students' learning progress, as illustrated in Figure 4, and includes planned learning time, total learning time, actual learning time, planned knowledge-learning points, and actual knowledge-learning points. This module also shows the test score and task score. These various design indicators are used to calculate SDL capabilities and performance. Moreover, the monitoring module provides visualization of progress toward the learning goals to help students reflect on their learning and plan their next steps, which can be seen as a stimulus for creative activity.

Figure 3. The screenshot of learning module

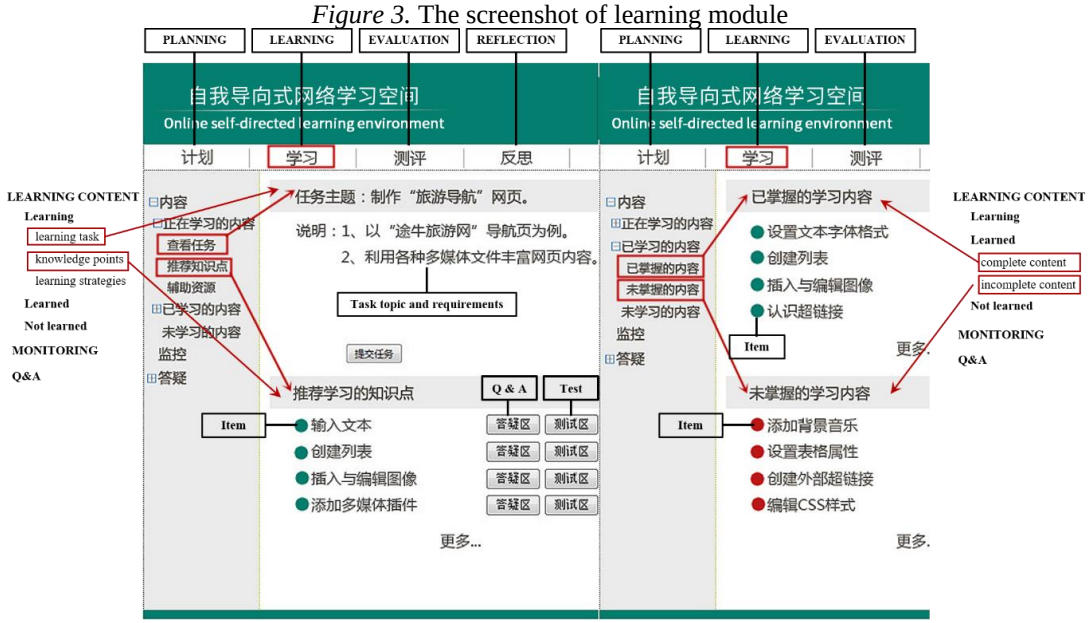
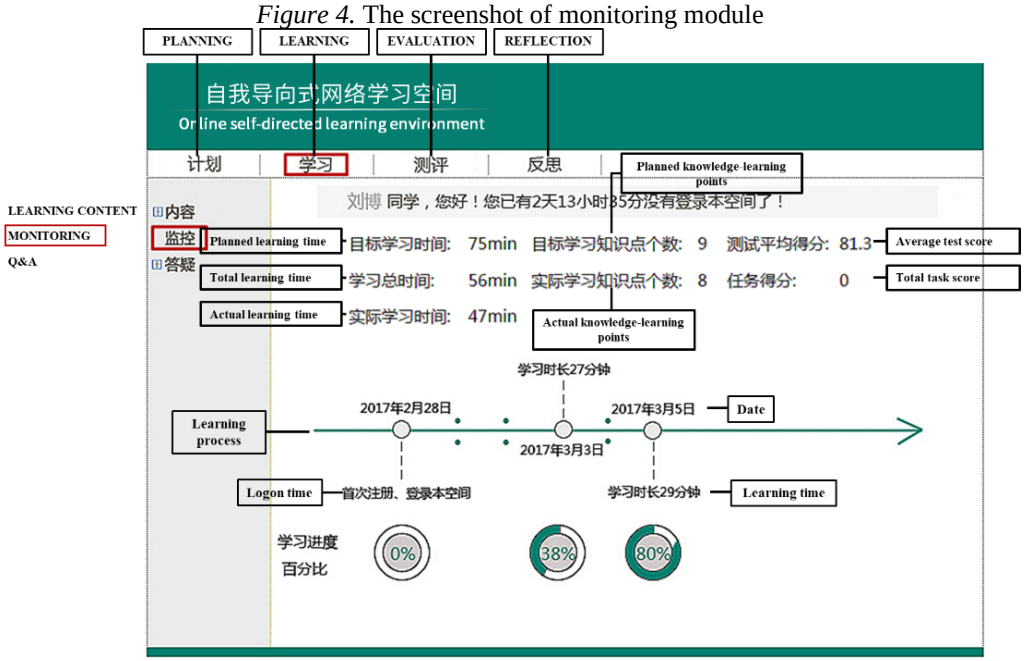


Figure 4. The screenshot of monitoring module



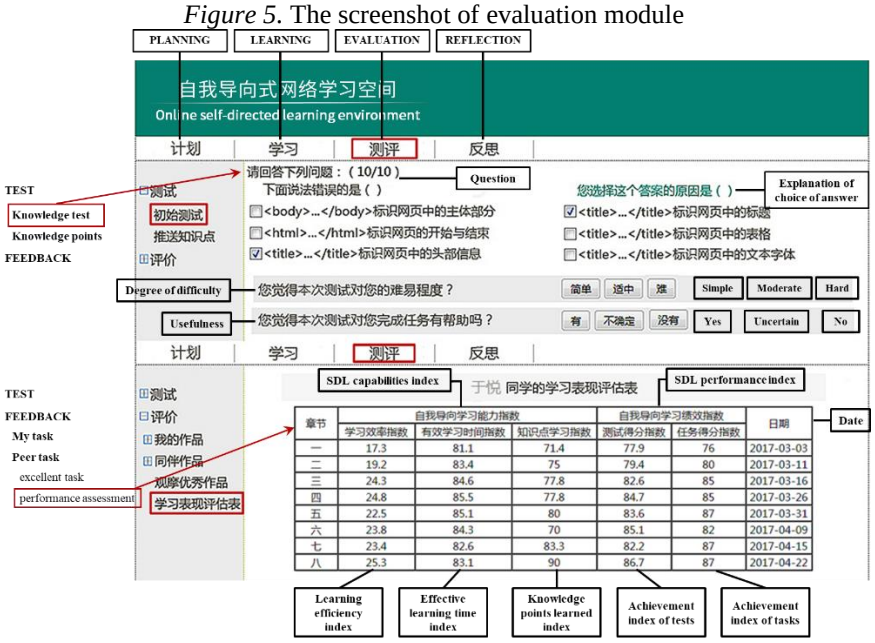
The Q&A module allows all students to share and discuss different opinions and thoughts. After the question is answered, and to reduce the uncertainty of peer knowledge, students can choose to close the question or keep the question open based on their judgment on the correctness of the answer. Meanwhile, the teacher can understand the primary problems by checking and answering students' queries. As students discuss the questions, creating a

coherent or compelling argument may compel them to identify gaps in their knowledge and to retrieve and modify their existing understanding with new ideas.

3.2.3. Evaluation

The evaluation module contains two parts: test and feedback. The two-tier test approach is used to identify students' learning status, as shown in Figure 5. A two-tier test consists of a set of multiple-choice items, including the question, answer choices, and the choice for the reasons (Yang et al., 2015). Subsequently, the misconception of knowledge is identified by both answers and reasons. Meanwhile, students who have used the OSDLE can diagnose their learning weaknesses and are enthusiastic about actively reflecting on the problem-solving process to facilitate creativity.

Figure 5. The screenshot of evaluation module



The feedback part of the OSDLE includes teacher feedback and peer feedback. In particular, the role of the course teacher and peers is essential for enhancing students' creativity. Students can view the teacher's comments while also observing samples of excellent work produced by their peers. This domain also produces a learning performance evaluation table, as shown in Figure 5, thereby enabling students to understand the gap between their performance and others and indicating the direction for further study.

3.2.4. Reflection

In the reflection module, students can articulate what they have learned during the task and accomplish the L part of the K-W-L chart. Because reflection does not develop automatically, it can be taught through effective facilitation (Ogle, 1986). Students can critique the skills and knowledge in the OSDLE and gain space for novel ideas and possibilities to emerge, which is a fundamental skill in encouraging creativity. At the same time, students can decide whether to share reflection logs. The reflection module promotes critical inquiry, engaged dialogue, and reflective practice (Song & Hill, 2007). Additionally, students review each other's online reflective journals in which they can reflect on their learning and bring forth creative ideas based on the habitual experience of the past.

4. Method

4.1. Participants

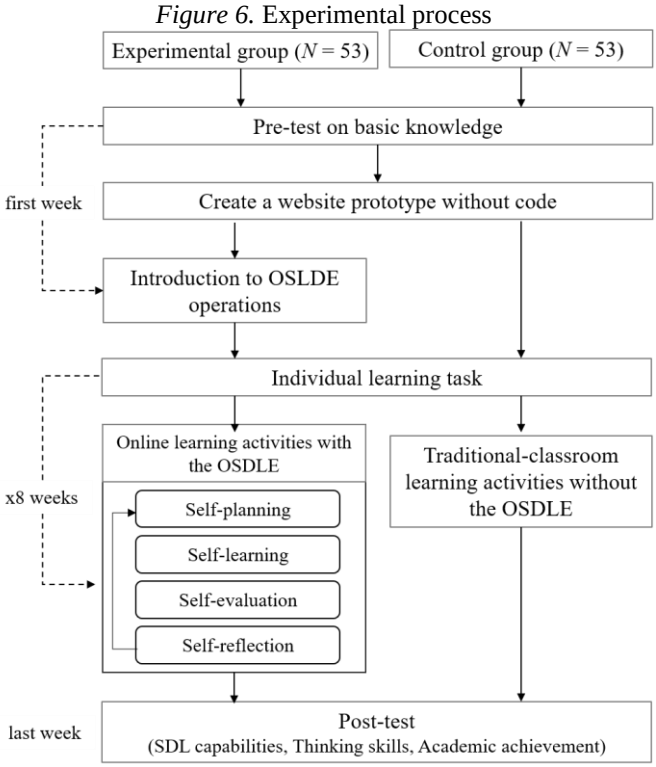
During the eight weeks of this study, 106 students at a university in Northeast China voluntarily participated. We want to mention that the original sample contained 106 university students who gave consent to use their data for research purposes. Two experiments on the course named web design and programming were conducted in two

separate semesters. In this respect, all participants were randomly assigned to the two groups. The experimental group (N = 53) was assigned to the experimental condition (i.e., using the OSDLE), and the control group (N = 53) to the control condition (i.e., not using the OSDLE).

4.2. Experimental procedure

As illustrated in Figure 6, the experimental design of study is introduced. The overall activity conditions were similar in experimental and control groups.

Figure 6. Experimental process



This study started with a set of prior knowledge tests as the pre-test to evaluate the participants' original understanding of web design and programming. The results of the pre-test indicated that there was no significant difference in prior academic achievement between the experimental group and control group (p = .89). At the beginning of the testing session, each participant was informed of the study's procedure and that their data would be handled confidentially. The participants were then asked to create a website prototype with basic design and functionality without writing any code before the experiment started. Later, the experimental group received an additional introduction session on the OSDLE and was instructed to use the functions in the four modules.

Next, the experimental and control groups received the same learning tasks and goals. The activity lasted eight weeks, with one task per week and a total of eight (i.e., Make a website with the body, font, br, and hr tags). The experimental group conducted learning using the OSDLE, while the control group received traditional teaching instruction in the classroom.

After the learning activities, the two groups of students were immediately administered the post-test during regular class time and were asked to submit a designed website with code based on what they had learned within one week. Finally, four experts were invited to assess the students' products to obtain data on students' thinking skills.

4.3. Measurement

The research measurement tools included logs of the supervised learning activities on the OSDLE, creative products, and pre- and post-tests.

To investigate the SDL capabilities of the experimental group during the experiment, their learning behaviors were recorded through the modules of OSDLE, including students' online time, task score, and the number of knowledge points learned (Chen, 2009). For online time, log data included the planned learning time to complete learning tasks, the total learning time that students were logged in the OSDLE, the learning time of each module, and the actual learning time, which is the total learning time minus the idle learning time (the time when the operation of the mouse or keyboard cannot be detected within a specific period). For the task score, each task was graded based on a scoring rubric that was designed on the basis of the course syllabus. The rubric was found to be consistently used by two professors and four researchers with good inter-rater reliability (Cronbach α = 0.92), as shown in Appendix Table 6. For the number of knowledge points learned, log data included the weekly knowledge-learning points completed.

The Consensual Assessment Technique (CAT) (Amabile, 1982) was used to assess the students' thinking skills in the pre- and post-experiments. Four independent experts blind-scored all products on a 5-point scale in originality, flexibility, and fluency, where the value of 5 represented the highest level of thinking skills. In this method, three major dimensions of the criteria for rating creative products were provided, which were attached in Appendix Table 7. Originality was determined by the percentage of pages that differed from the categories covered by the reference sample. Flexibility represented an estimate of the degree of website layout friendliness. Fluency represented an estimate of how many types of functions were designed. To prevent the order effect, each judge rated the compositions in a different random order. The four experts' ratings on the three dimensions of originality, flexibility and fluency were then averaged separately to produce the students' scores on each dimension, and the average of the three dimensions was used as a measure of thinking skills for participants. To access internal consistency of the CAT's dimensions, Cronbach's alpha calculated for originality was .84, for flexibility was .80 and for fluency was .82, indicating high consistency.

To investigate the differences in participants' academic achievement, the scores of pre- and post-tests were analyzed. Both the pre- and post-tests consisted of 20 two-tier multiple-choice questions, with a total score of 100. A two-tiered question is considered correct only if both tiers were answered correctly. These tests were focused on the content of the lessons (i.e., the website design) with the same knowledge but different levels of difficulty. The pre-test with relatively low difficulty values was given to assess the students' pre-performance before the experiment, and the post-test score reflected the students' post-performance.

5. Results

IBM SPSS was applied to analyze the creativity performance of the participants, including the results of SDL capabilities, thinking skills, and academic achievement.

5.1. Self-directed learning capabilities

The SDL capabilities index includes the learning efficiency index (T1), effective learning time index (T2), and knowledge points learned index (K).

The learning efficiency index refers to the efficiency of students completing a task in the OSDLE. The formula for calculating the learning efficiency index is shown in Table 1, where t-time(s) is the s{th}student's total learning time, and p-time(s) is the planned learning time set by the s{th} student in the planning module.

The effective learning time index is the ratio between actual and total learning times. The formula for calculating the effective learning time index is shown in Table 1, where a-time(s) is the s{th} student's actual learning time.

The knowledge points learned index is evaluated based on the ratio of the knowledge-learning points passed by the student in the weekly tests and the planned knowledge-learning points set by the students. The knowledge points were different every week because of different contents. The larger the value, the greater the number of knowledge point learned by the student. The formula for calculating the knowledge points learned index is shown in Table 1, where a-knowledge(s) is the amount of actual knowledge-learning points of the s{th} student during the learning process, and p-knowledge(s) is the amount of planned knowledge-learning points set by the s{th} student. When the knowledge points learned index is greater than or equal to 1, it is obtained by 1.

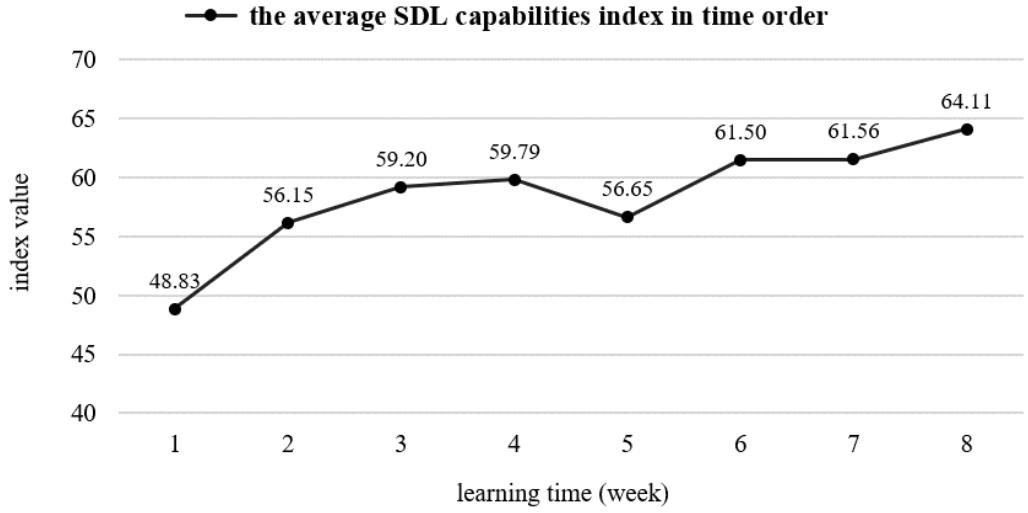
By analyzing the mean value of the SDL capabilities sub-index of 53 students in the experimental group, it can be seen that the SDL capabilities index of the experimental group showed a general trend of gradual

improvement within eight weeks. A significant difference was observed in the experimental group's SDL capabilities to complete the first and eighth tasks (p = .005 < .05). More details of the students' statistical information of SDL capabilities are listed in Appendix Table 8. In addition to a slight decrease in the fifth week, the interview with the students in the subsequent period revealed that the fifth week was during the midterm exam, which might cause students to reduce their engagement. Therefore, the results revealed that the OSDLE had a positive effect on improving students' SDL capabilities (the average of the learning efficiency index, effective learning time index, and knowledge points learned index), as shown in Figure 7.

Table 1. Formula for SDL capabilities index

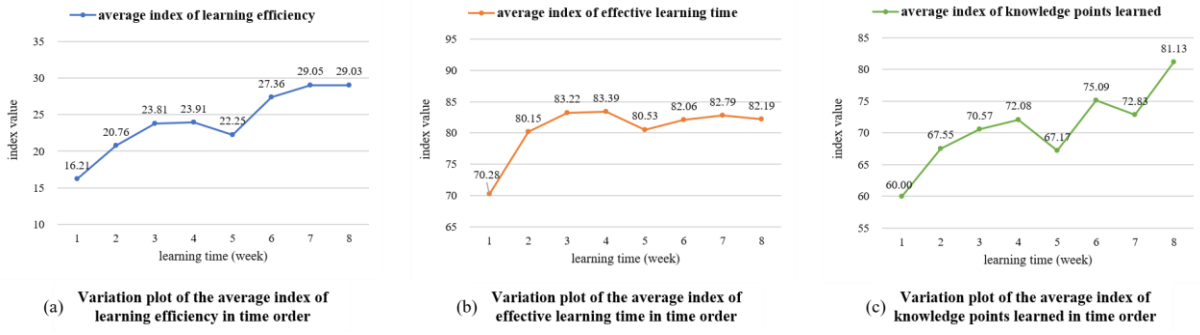
SDL capabilities sub-indicator	Formula
Learning efficiency index(T1)	$T1 = \frac{\min[t-time(s), p-time(s)]}{p-time(s)} * 100$
Effective learning time index(T2)	$T2 = \frac{a-time(s)}{t-time(s)} * 100$
Knowledge points learned index(K)	$K = \left(\frac{a-knowledge(s)}{p-knowledge(s)}\right) * 100, if(K \geq 1),\ then\ K = 1$

Figure 7. Variation plot of the average SDL capabilities index in time order



The learning efficiency index continued to increase gradually, indicating that students could gradually manage their learning time effectively within eight weeks, as shown in Figure 8(a). The growth rate of the effective learning time index increased significantly during the first three weeks but had a slower growth rate over time. This may be attributed to the fact that, although varied and novel learning activities could initially spark a high level of willpower and engagement, they did not encourage perseverance. As shown in Figure 8(b), the effective learning time index depending on students' effective engagement fluctuated greatly, and heavy workloads in the fifth week could negatively affect students' engagement. The knowledge points learned index generally showed an upward trend during the experiment, indicating that students became more efficient in achieving self-planned learning goals, as shown in Figure 8(c).

Figure 8. Variation plot of the average SDL capabilities sub-indicator in time order



(a) Variation plot of the average index of learning efficiency in time order

(b) Variation plot of the average index of effective learning time in time order

(c) Variation plot of the average index of knowledge points learned in time order

5.2. Thinking skills

The scores for the students' website design products were used to evaluate thinking skills, applying originality, flexibility, and fluency as three components to confirm the effect of the OSDLE on the final product designs. The principal results of the thinking skills of the control and experimental groups are shown in Table 2.

Table 2. Descriptive statistics for students' thinking skills

Item	Before the experiment				After the experiment			
	Control group		Experimental group		Control group		Experimental group	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Originality	2.77	1.28	2.88	1.52	2.98	1.30	3.64	1.03
Flexibility	2.96	1.28	2.88	1.40	3.11	1.17	3.96	1.05
Fluency	2.94	1.30	3.22	1.50	2.94	1.26	3.56	1.10
Total score	2.89	0.99	3.00	1.06	3.01	0.68	3.72	0.66
p	.594				.000 ***			

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

Table 2 shows that no significant difference in students' thinking skills was observed between the experimental and the control groups before the experiment ($p = .594$). After the experiment, the results revealed a significant difference in thinking skills between the experimental and control groups ($p < .001$). Moreover, the average gain score of students' thinking skills in the experimental group ($M = 3.72$, $SD = .66$) was significantly greater than that in the control group ($M = 3.01$, $SD = .68$). Therefore, the OSDLE had a beneficial effect in increasing the level of thinking skills.

5.3. Academic achievement

The study also investigated the impacts of the OSDLE on students' academic achievement. Table 3 shows the summary statistics of the t-test. The t-test showed that there was a statistical difference in academic achievement between the experimental group and the control group ($p = .045$) after the learning activities. Additionally, the average gain score of students' academic achievement in the experimental group ($M = 89.03$, $SD = 1.20$) was significantly higher than that in the control group ($M = 85.27$, $SD = 1.41$). The results indicated that students who studied in the OSDLE performed significantly better than those who studied in the traditional classroom environment.

Table 3. Descriptive statistics for students' academic achievement

Group	*N*	*M*	*SD*	*t*	*df*	*p*	*MD*	*SE* difference
Control group	53	85.27	1.41	2.03	104	.045*	3.75	1.85
Experimental group	53	89.03	1.20					

Note. *$p < .05$.

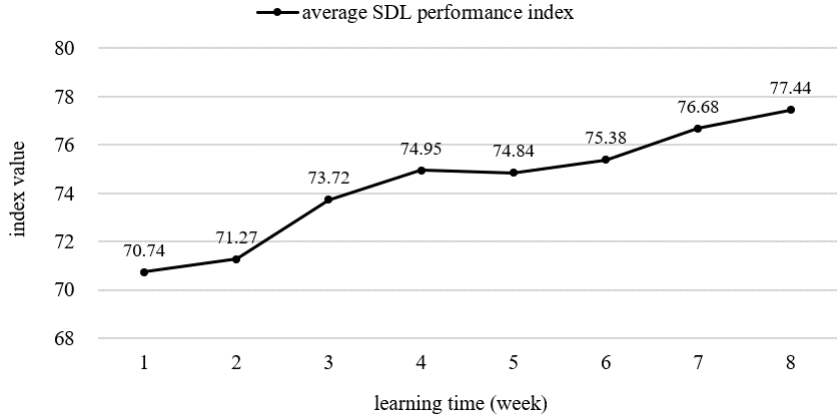
Subsequently, to further analyze the learning process of the experimental group, we calculated their SDL performance index in the OSDLE. The SDL performance index is the mean value of the tests score index (A1) and tasks score index (A2). The formula definition is shown in Table 4, where n is the total number of passed tests, s-test(s)i is the s{th} student's actual test score for the i{th} test, t-testi is the target test score of the i{th} test set by the teacher, s-task(s) is the s{th} student's actual task score, and t-task(s) is the target task score set by the teacher.

Table 4. Formula for SDL performance index

SDL performance sub-indicator	Formula
Average achievement index of tests (A1)	$A1 = \left(\frac{\sum_{i=1}^{n} s-test(s)_i}{\sum_{i=1}^{n} t-test_i}\right) * 100$
Average achievement index of tasks (A2)	$A2 = \frac{s-task(s)}{t-task(s)} * 100$

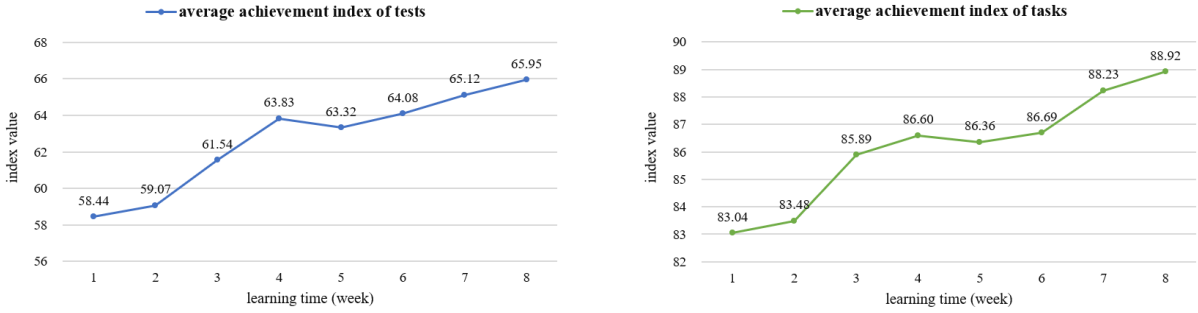
The SDL performance index curve shows that the overall performance of the experimental group indicated an upward trend during the eight-week learning period, as shown in Figure 9. More details of the students' statistical information of SDL performance are shown in Appendix Table 9. A significant difference in SDL performance was observed between the first and eighth tasks ($p = .002 < .05$).

Figure 9. Variation plot of the average SDL performance index of the experimental group in time order



By analyzing the tests score index and tasks score index of 53 students in the experimental group, it can be seen that the increase was particularly noteworthy in the first four weeks, while the students' SDL performance index changed unstably in the fifth week, as shown in Figure 10. The main reason could be that students could skillfully use the OSDLE platform after one week, and their SDL performance was also enhanced. However, during the midterm exam, the learning process was suppressed under the influence of excessive academic burden, which had a negative impact on students' SDL performance. Seven weeks later, the growth rate of SDL performance slowed down, but on the whole, it maintained a slight upward tendency. This may be explained by the fact that students had almost fully adapted to study in the OSDLE, and their performance tended to improve steadily.

Figure 10. Variation plot of the average SDL performance sub-indicator in time order



(a) **Variation plot of the average achievement index of tests in time order**

(b) **Variation plot of the average achievement index of tasks in time order**

6. Discussion and conclusions

To improve students' creativity performance, this study proposed the OSDLE based on the conceptual framework of SDL. This study introduced the OSDLE to a university course to determine whether the OSDLE would positively affect students' creativity performance by analyzing students' SDL capabilities, thinking skills, and academic achievement. The results of the quasi-experiment indicated that the proposed OSDLE significantly improved students' SDL capabilities, thinking skills, and academic achievement. It suggests that using the OSDLE could better support students' creativity performance. The following session discusses the research questions presented at the beginning of the paper.

OSDLE - SDL capabilities (RQ.1). An interesting finding is that the varying curve of SDL capabilities index indicated that the students' SDL capabilities continued to improve gradually in the process of using the OSDLE, and a significant difference could be observed between the first task and the eighth task. Although SDL capabilities have continued to increase in the initial stage, the ascent rate decreased as time. The reason might be that students' SDL capabilities are related to learning motivation. Li et al. (2021) showed a significant relationship between SDL capabilities and learning motivation levels. After the initial enthusiasm, students might feel overwhelmed by various learning activities like the upcoming exams, affecting learning motivation and further impacting SDL capabilities.

With this in mind, future research must consider the maintenance of student learning motivation to improve SDL capabilities. On the one hand, it is necessary to facilitate SDL capabilities-based adaptive feedback systems for students (Jiang et al., 2019), which could enable them to engage more in the planning and monitoring interactions in the OSDLE, and then subsequently affect their learning motivation. On the other hand, the role of self-efficacy in creativity promotion cannot be ignored. Schweder (2019) found that self-effective students tended to be more motivated in learning and thus had higher SDL capabilities. Accordingly, it is recommended that students studying in the OSDLE should be given sufficient guidance and training on goal-setting and monitoring to maintain pleasure during the learning process, which would keep them motivated.

OSDLE - Thinking skills (RQ.2). The findings agree with the positive relationship between the specific learning environment and thinking skills, which is confirmed by the results of Gralewski and Karwowski (2019). OSDLE can significantly improve students' thinking skills statistically. While the course instruction does not emphasize systematic cultivation of thinking skills, OSDLE guides students to make clear and specific learning plans for themselves and drives them to engage in a higher level of task commitment spontaneously. On the whole, OSDLE can exert a significant effect on the development of students' thinking skills.

Seen from the perspective of this paper, OSDLE is related to creativity through its connection to thinking skills. Accordingly, creativity performance can become advanced by developing thinking skills based on the analysis of learning tasks (Rhodes, 1961). Therefore, students' active participation in the course should be encouraged. In this experiment, students participated in the process of creating the website as a creative product. Integrating an appropriate level of knowledge and experience could help students stimulate effective retrieval of knowledge and break the inherent thinking framework for generating original ideas (Ness, 2012), ultimately achieving a high level of thinking skills. This step also facilitates the attainment of creativity performance.

OSDLE - Academic achievement (RQ.3). The experiment results also revealed that the students in OSDLE demonstrated significantly higher academic achievement than those in the traditional classroom. In addition, the SDL performance of students in the experimental group maintained an upward trend. Therefore, the findings indicated the effectiveness of the OSDLE in improving academic achievement. These results are supported by Dunn and Kennedy (2019), whose study investigated the associations between technologies and academic achievement.

Although students are digital natives, individuals are prone to technical pressure because of the failure in timely responding to changes brought by new technologies. Therefore, elaboration prompts are recommended for OSDLE to decrease learning inefficiency caused by system complexity in the future. Before students start learning, it is suggested that a conference on the platform operation should be organized, where the user function can be introduced (Huang et al., 2017). In particular, to promote a creative learning experience, students need to build on their ideas as the first step in developing their creative capacity (Hwang et al., 2021). When students are proficient in using the OSDLE, they can independently participate in learning and interaction through the platform, thereby promoting creativity.

Overall, this study introduced the OSDLE with the functions of planning, learning, evaluation, and reflection into a university course, and examined the effects of the OSDLE on creativity performance. The findings revealed that the OSDLE significantly improved SDL capabilities, thinking skills, and academic achievement. It demonstrated that using the OSDLE could promote students' creativity performance. The main contribution of this study has implications for researchers studying creativity. This study highlighted the contribution of OSDLE to promote creativity performance. OSDLE enables students to develop learning capabilities and thinking skills, enhance academic achievement, and become independent learners to actively generate various original ideas that can foster their creativity in return. Therefore, OSDLE has become one of the valuable environments to support creativity. In addition, this study contributes to the growing body of literature on improving creativity. First, the present study broadened the understanding of the relationship between creativity performance and SDL through pedagogical practice. Second, this study provided the OSDLE to examine the support of educational technology for creativity development in authentic contexts. Finally, some practical implications can be provided for instructors. Specifically, to promote students' creativity performance, proper and sufficient scaffolding should be provided for them, as the OSDLE is a way of planning the learning process by students' choices and pace.

The main limitation of this study is that the research was carried out in a single course. In future work, it would be advisable to implement practices in a greater number of courses from different disciplines, as we did in some of our experiments. Another limitation is that the research mainly focused on the computer-based environment but did not consider a ubiquitous learning environment. Further studies are needed to apply the results to mobile learning to increase the practical value of this research.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (NSFC) under Grant 62077012.

References

Amabile, T. M. (1982). Social psychology of creativity: A Consensual assessment technique. *Journal of Personality & Social Psychology*, *43*(5), 997-1013. https://doi.org/10.1037/0022-3514.43.5.997

Bloom, B.S. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain*. David McKay.

Candy, P. C. (1991). *Self-direction for lifelong learning: A Comprehensive guide to theory and practice*. Jossey-Bass.

Chen, C. -M. (2009). Personalized e-learning system with self-regulated learning assisted mechanisms for promoting learning performance. *Expert Systems with Applications*, *36*(5), 8816-8829. https://doi.org/10.1016/j.eswa.2008.11.026

Chiu, F. -C., & Tu, P. L. -P. (2014). The Priming effect of military service on creativity performance. *Psychological Reports*, *114*(2), 509-527. https://doi.org/10.2466/21.PR0.114k22w1

Chu, R. J., Chu, A. Z., Weng, C., Tsai, C. -C., & Lin, C. -C. (2012). Transformation for adults in an internet-based learning environment — Is it necessary to be self-directed? *British Journal of Educational Technology*, *43*(2), 205-216. https://doi.org/10.1111/j.1467-8535.2010.01166.x

Dewey, J. (1916). *Democracy and education: An Introduction to the philosophy of education*. The Macmillan Company.

Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge University Press.

Dunn, T. J., & Kennedy, M. (2019). Technology Enhanced Learning in higher education; motivations, engagement and academic achievement. *Computers & Education*, *137*, 104-113. https://doi.org/10.1016/j.compedu.2019.04.004

Garrison, D. R. (1997). Self-directed learning: Toward a comprehensive model. *Adult Education Quarterly*, *48*(1), 18-33. https://doi.org/10.1177/074171369704800103

Gralewski, J., & Karwowski, M. (2019). Are teachers' ratings of students' creativity related to students' divergent thinking? A meta-analysis. *Thinking Skills and Creativity*, *33*, 100583. https://doi.org/10.1016/j.tsc.2019.100583

Guglielmino, L. M. (1977). *Development of the self-directed learning readiness scale*. University of Georgia.

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *The Journal of Creative Behavior*, *1*(1), 3-14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Hardy, J. H., Ness, A. M., & Mecca, J. (2017). Outside the box: Epistemic curiosity as a predictor of creative problem solving and creative performance. *Personality and Individual Differences*, *104*, 230-237. https://doi.org/10.1016/j.paid.2016.08.004

Hernández-Torrano, D., & Ibrayeva, L. (2020). Creativity and education: A Bibliometric mapping of the research literature (1975-2019). *Thinking Skills and Creativity*, *35*, 100625. https://doi.org/10.1016/j.tsc.2019.100625

Holland, P. W., & Hoskens, M. (2003). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika*, *68*, 123-149. https://doi.org/10.1007/BF02296657

Hong, O., & Song, J. (2020). A Componential model of Science Classroom Creativity (SCC) for understanding collective creativity in the science classroom. *Thinking Skills and Creativity*, *37*, 100698. https://doi.org/10.1016/j.tsc.2020.100698

Huang, Y. -M., Shadiev, R., Sun, A., Hwang, W. -Y., & Liu, T. -Y. (2017). A Study of the cognitive diffusion model: Facilitating students' high level cognitive processes with authentic support. *Educational Technology Research and Development*, *65*(3), 505-531. https://doi.org/10.1007/s11423-016-9475-0

Hwang, W. -Y., Hariyanti, U., Chen, N. -S., & Purba S. W. D. (2021). Developing and validating an authentic contextual learning framework: Promoting healthy learning through learning by applying. *Interactive Learning Environments*, 1-13. https://doi.org/10.1080/10494820.2021.1876737

Isaksen, S. G., Puccio, G. J., & Treffinger, D. J. (1993). An Ecological approach to creativity research: Profiling for creative problem solving. *The Journal of Creative Behavior*, *27*(3), 149-170. https://doi.org/10.1002/j.2162-6057.1993.tb00704.x

Jiang, Q., Li, Y., Sun, J., Zhao, W., & Liu, H. X. (2019). Adaptive blended MOOC: New design paradigm of MOOC. *China Educational Technology*, (9), 82-90.

Kim, D., Jung, E., Yoon, M., Chang, Y., Park, S., Kim, D., & Demir, F. (2021). Exploring the structural relationships between course design factors, learner commitment, self-directed learning, and intentions for further learning in a self-paced MOOC. *Computers & Education*, *166*(6), 104171. https://doi.org/10.1016/j.compedu.2021.104171

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1-48. https://doi.org/10.1207/s15516709cog1201_1

Knowles, M. (1975). *Self-directed learning: A Guide for learners and teachers*. Association Press.

Lee, Z. D. (2019). A Study on the level of creative thinking ability and influence factors of high school self-directed learning entrance admission students. *Journal of Education & Culture*, *25*(2), 529-548. https://doi.org/10.24159/joec.2019.25.2.529

Lemmetty, S., & Collin, K. (2021). Self-directed learning in creative activity: An Ethnographic study in technology-based work. *Journal of Creative Behavior*, *55*(1), 105-119. https://doi.org/10.1002/jocb.438

Li, H., Majumdar, R., Chen, A. M.-R., & Ogata, H. (2021). Goal-oriented active learning (GOAL) system to promote reading engagement, self-directed learning behavior, and motivation in extensive reading. *Computers & Education*, *171*, 104239. https://doi.org/10.1016/j.compedu.2021.104239

Ma, H.-H. (2009). The Effect size of variables associated with creativity: A Meta-analysis. *Creativity Research Journal*, *21*(1), 30-42. https://doi.org/10.1080/10400410802633400

Mamun, M. A. A., Lawrie, G., & Wright, T. (2020). Instructional design of scaffolded online learning modules for self-directed and inquiry-based learning environments. *Computers & Education*, *144*, 103695. https://doi.org/10.1016/j.compedu.2019.103695

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

Mishra, P., Fahnoe, C., Henriksen, D., & the Deep-Play Research Group. (2013). Creativity, self-directed learning and the architecture of technology rich environments. *Techtrends*, *57*(1), 10-13. https://doi.org/10.1007/s11528-012-0623-z

Morris, T. H. (2020). Creativity through self-directed learning: Three distinct dimensions of teacher support. *International Journal of Lifelong Education*, *39*(2), 168-178. https://doi.org/10.1080/02601370.2020.1727577

Morris, T. H., & Rohs, M. (2021). Digitization bolstering self-directed learning for information literate adults — A Systematic review. *Computers and Education Open*, *2*, 100048. https://doi.org/10.1016/j.caeo.2021.100048

Ness, R. B. (2012). *Innovation generation: How to produce creative and useful scientific ideas*. Oxford University Press.

Ogle, D. M. (1986). K-W-L: A Teaching model that develops active reading of expository text. *The Reading Teacher*, *39*(6), 564-570. https://doi.org/10.1598/RT.39.6.11

Rhodes, M. (1961). An Analysis of creativity. *Phi Delta Kappan*, *42*(7), 305-310. http://www.jstor.org/stable/20342603

Rubenstein, L. D., Callan, G. L., & Ridgley, L. M. (2018). Anchoring the creative process within a self-regulated learning framework: Inspiring assessment methods and future research. *Educational Psychology Review*, *30*(3), 921-945. https://doi.org/10.1007/s10648-017-9431-5

Schweder, S. (2019) The Role of control strategies, self-efficacy, and learning behavior in self-directed learning, *International Journal of School & Educational Psychology*, *7*(1), 29-41. https://doi.org/10.1080/21683603.2018.1459991

Shadiev, R., Hwang, W. -Y., & Liu, T. -Y. (2018). Investigating the effectiveness of a learning activity supported by a mobile multimedia learning system to enhance autonomous EFL learning in authentic contexts. *Educational Technology Research and Development*, *66*(4), 893-912. https://doi.org/10.1007/s11423-018-9590-1

Smith, M. S. (2009). Opening education. *Science*, *323*(5910), 89-93. https://doi.org/10.1126/science.1168018

Song, L., & Hill, J. R. (2007). A Conceptual model for understanding self-directed learning in online environments. *Journal of Interactive Online Learning*, *6*(1), 27-42. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.109.5170

Sternberg, R. J. (2006). The Nature of creativity. *Creativity Research Journal*, *18*(1), 87-98. https://doi.org/10.1207/s15326934crj1801_10

Tang, T., Vezzani, V., & Eriksson, V. (2020). Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity*, *37*, 100696. https://doi.org/10.1016/j.tsc.2020.100696

Tekkol, I. A., & Demirel, M. (2018). An Investigation of self-directed learning skills of undergraduate students. *Frontiers in Psychology*, *9*, 2324. https://doi.org/10.3389/fpsyg.2018.02324

Toh, W., & Kirschner, D. (2020). Self-directed learning in video games, affordances and pedagogical implications for teaching and learning. *Computers & Education*, *154*, 103912. https://doi.org/10.1016/j.compedu.2020.103912

Torrance, E. P., & Mourad, S. (1978). Some creativity and style of learning and thinking correlates of Guglielmino's self-directed learning readiness scale. *Psychological Reports*, *43*(3), 1167-1171. https://doi.org/10.2466/pr0.1978.43.3f.1167

Yang, T.-C., Hwang, G.-J., Yang, S. J. H., & Hwang, G.-H. (2015). A Two-tier test-based approach to improving students' computer-programming skills in a web-based learning environment. *Educational Technology & Society*, *18*(1), 198-210.

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

Appendix

Figure 11. An example of two-tier multiple-choice items for tests

Table 5. Illustrative example of a two-tier test item

Question:	
In the following example, the basic structure of an HTML5 document consists of 4 elements: 1.<html> 2.<head> 3.<title> TITLE</title> 4.</head> 5.<body> 6. CONTENT 7.</body> 8.</html> Which of the following statements is incorrect for this basic structure?	
First tier	Second tier
a. <body> … </body> represents the content of an HTML document	(a1) The <body> element defines the document's body. (a2) The <body> element represents introductory content. (a3) The <body> element contains meta information about an HTML page.
b. <html>…</html> delimits the beginning and the end of an HTML document	(b1) The <html> element is the root element of an HTML page. (b2) The <html> element defines that this document is an HTML5 document. (b3) The <html> element defines a paragraph.
c. <title>…</title> defines the head of an HTML document	(c1) The <title> element defines a large heading of an HTML page. (c2) The <title> element defines a chat heading. (c3) The <title> element defines the document's title that is shown in a browser's title bar or a page's tab.

Table 6. Task rubric

Criteria	(Strongest) 5	4	3	2	(Weakest) 1	Percentage
Completeness & Accuracy	The product is complete and correct.	The product is complete, but there are still some	The product is almost complete, but half of the	The product is somehow complete, but most of the	The product is not complete or unrelated.	40%

		mistakes in the content.	details are wrong.	details are wrong.		
Layout	The product has an exceptionally attractive and easily navigable layout.	The product has an attractive and usable layout.	The product has a usable layout, but some parts may appear busy or boring.	The product has a cluttered or confusing layout.	The product has an unusable and disorganized layout.	20%
Design	The project has five of the following: • Captures attention • Visually interesting • Engaging • Well crafted • Has an aesthetic quality	The project has four of the following: • Captures attention • Visually interesting • Engaging • Well crafted • Has an aesthetic quality	The project has three of the following: • Captures attention • Visually interesting • Engaging • Well crafted • Has an aesthetic quality	The project has two of the following: • Captures attention • Visually interesting • Engaging • Well crafted • Has an aesthetic quality	The project has zero or one of the following: • Captures attention • Visually interesting • Engaging • Well crafted • Has an aesthetic quality	20%
Creativity	Was extremely clever and presented with originality. A unique approach that truly enhanced the product.	Thoughtfully and uniquely presented. Was clever at times.	Added some original touches to enhance the product, but did not incorporate them throughout.	Have only a few unique aspects. Most elements are copied from the sample.	Unoriginal or borrowed product.	20%

Table 7. The Creative Product Scale

Please use your subjective opinion of three dimensions of creativity to evaluate each creative product individually. The description provided is only a suggestion to guide your evaluation.

1. Originality: Refers to the percentage of pages that differed from the categories covered by the reference sample.
1- Generates repeated ideas.
2- Generates a few unique or unusual ideas.
3- Generates several unique or unusual ideas.
4- Generates a sufficient volume of unique or unusual ideas.
5- Takes a novel, unique or unusual approach to idea generation.
2. Flexibility: Refers to an estimate of the degree of website layout friendliness.
1- Presents ideas in isolation.
2- Simple connections are made between a part of ideas.
3- Reasonable connections are made between ideas.
4- Often makes effective connections between ideas using various organizational techniques.
5- Makes precise and complex connections between different related ideas in unexpected ways.
3. Fluency: Refers to an estimate of how many types of functions were designed.
1- Shows an inability to design any functions creatively.
2- Presents functions that are vague or incomplete and are not considered to be unique.
3- Presents a few functions that are considered to be somewhat valuable and unique.
4- Presents sufficient functions to be considered valuable and unique.
5- Shows an impressive level of creative, diverse functional design.

Table 8. The students' statistical information of eight weeks based on the average learning efficiency index, the average effective learning time index, the average knowledge points learned index, and the average SDL capabilities index

Learning time (week)	Number of students	Learning efficiency index		Effective learning time index		Knowledge points learned index		SDL capabilities index	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	53	16.21	2.45	70.28	3.59	60.00	3.23	48.83	1.89
2	53	20.76	3.29	80.15	3.73	67.55	3.62	56.15	1.89
3	53	23.81	3.75	83.22	2.90	70.57	3.71	59.20	2.05
4	53	23.91	4.13	83.39	3.63	72.08	4.15	59.79	2.60
5	53	22.25	3.68	80.53	3.34	67.17	3.00	56.65	1.79
6	53	27.36	4.14	82.06	3.21	75.09	3.84	61.50	2.49
7	53	29.05	4.07	82.79	3.45	72.83	3.57	61.56	2.42
8	53	29.03	3.97	82.19	2.91	81.13	2.71	64.11	1.89

Table 9. The students' statistical information of eight weeks based on the average achievement index of tests, the average achievement index of tasks, and the average SDL performance index

Learning time (week)	Number of students	Achievement index of tests		Achievement index of tasks		SDL performance index	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
1	53	58.44	2.29	83.04	1.76	70.74	1.45
2	53	59.07	3.07	83.48	2.11	71.27	1.91
3	53	61.54	2.963	85.89	1.94	73.72	1.75
4	53	63.83	3.02	86.60	1.88	74.95	1.73
5	53	63.32	3.46	86.36	1.76	74.84	1.63
6	53	64.08	3.33	86.69	2.29	75.38	2.05
7	53	65.12	3.25	88.23	1.67	76.68	1.99
8	53	65.95	2.58	88.92	1.26	77.44	1.26

Creative Situated Augmented Reality Learning for Astronomy Curricula

Chia-Chen Chen[1], Hong-Ren Chen[2*] and Ting-Yu Wang[1]

[1]Department of Management Information Systems, National Chung Hsing University, Taiwan // [2]Department of Digital Content and Technology, National Taichung University of Education, Taiwan // emily@nchu.edu.tw // hrchen@mail.ntcu.edu.tw // a7278868@gmail.com

[*]Corresponding author

ABSTRACT: Many elementary school students find astronomical knowledge difficult to attain. Students cannot observe planetary motion in the universe, which makes the construction of astronomical knowledge abstract and incomprehensible for many students. To cope with this dilemma, this study proposed creative situated learning via augmented reality (AR) and developed an AR-based Cosmos Planet Go App to simulate the motion of planets in the universe. This allowed students to understand the characteristics and features of each planet through its simulated motion in the universe. This study adopted a quasi-experimental method and the qualitative analysis to conduct experiments on teaching astronomy in an elementary school in central Taiwan. The control group students were taught using traditional classroom narrative teaching, and the experimental group students were taught using the AR-based Cosmos Planet Go App. The results showed that students who learned with the use of the AR-based Cosmos Planet Go App performed significantly better than the control students on measures of learning effectiveness, learning motivation, and flow experience. Moreover, learning engagement, which occurs when students can use multiple perspectives to solve problems, is the most important element for evaluating the AR-learning environment in creative situations. This study extended the research field of digital technology-assisted learning to the discussion of integrated creative learning environment, which can be used as the basis and reference for scholars' research.

Keywords: Creative situated learning, ARCS motivation model, Augmented reality, Astronomy curriculum

1. Introduction

The concept of situated learning suggests that learning is not merely a result of knowledge imparted through teaching, and that the process of learning is also a source of knowledge. Situated learning theory emphasizes that learning occurs in the situated process of constructing knowledge. Appropriate situated learning contributes to the improvement of teaching effectiveness. From the perspective of the learner, the learning situation can be used to combine life experience with the current situation to produce in-depth knowledge that can be deeply understood and meaningful. If the knowledge learned is separated from the situation, it becomes a cluster of abstract symbols with limited meaning. At present, many studies use creative situated learning to place students into, and have them interact with, the teaching situation, actively exploring and reflecting during the interaction so that they can construct their own knowledge and ability (Chen & Lin, 2016b; Hwang et al., 2020).

A creative learning environment that simulates a real situation through digital technology can incentivize students to be more involved in gaining knowledge, effectively encouraging them to create different ideas and extend the application of the things they learn. Lau (2011) examined the key role of a creative learning environment in design education, focusing on the actual creative work and the learning space. That study showed that virtual reality/augmented reality could provide a learning environment that highly simulated real situations. An imaginative learning space can help students to enjoy a game-like learning environment and find creative fun in the learning process.

Richardson and Mishra (2018) proposed the SCALE creative learning framework to promote understanding and extension of learned knowledge and to enhance its application. This framework for a creative learning environment has been used to guide educators in supporting the construction and development of creative learning environments. The learning environment affecting students' creativity includes the classroom environment, learner participation, and the classroom learning climate. Activities that support creative learning include exploration of new media technologies, engagement in fantasy games, and digital model production and design (David et al., 2013). The creative learning environment does not focus on creating a visually strange or eye-catching learning environment, rather, it helps students to extend the requirements of knowledge through the creative learning environment (Stolaki & Economides, 2018).

Thanks to recent developments in science and technology, scientists have revealed the mysteries of the universe through various space detectors, and the software and hardware for simulating astronomical phenomena have also brought forth many innovative ideas. At present, part of the astronomy curriculum focuses on observing the

starry sky, but students' observations may be limited due to the inconvenience of time and place. This problem can be addressed through the aid of technology (Zhang et al., 2014). Some scholars have also noted that students may use augmented reality (AR) devices to carry out authentic learning activities, which can help students to explore the real-world environment, enable them to have a more authentic experience, and enhance their understanding of learning content (Chin et al., 2018a; Chin et al., 2018b; Tseng et al., 2016). Thus, the application of appropriate assisted learning technologies can help students to overcome learning obstacles and make them more willing to understand and learn astronomy without being influenced by these external factors. Li and Keller (2018) argued that most studies have only used scales to measure students' learning motivation and that curricula were actually not designed based on ARCS-based teaching strategies. Therefore, this study used the ARCS Motivation Model and its associated teaching strategies to design a curriculum, and it used the ARCS motivation scale to measure students' learning motivation. Flow experience represents an individual's state of mind. When a person has a flow experience, he/she experiences reality as positive, beneficial, controllable, and challenging and then concentrates on the things going on (Bressler & Bodzin, 2013; Hsu, 2017). This study measured the students' flow experience state to understand whether they were immersed in their learning.

When teaching astronomy, teachers usually face the problem that students can only imagine planetary motion in the universe from textbooks and cannot observe it in reality. This makes it difficult for students to understand the relevant concepts, resulting in poor learning performance and low learning motivation. The connections among the orbits of the planets in the universe as a whole may not be understood; additionally, students may not be motivated to engage in creative performative learning. To address this difficulty in teaching astronomical knowledge, this study proposed developing an AR-based Cosmos Planet Go APP by using creative, situated AR learning, which applies AR technology to simulate the reality of planetary motion in the universe. With this tool, students can discuss the learning content and engage in creative expression around different ideas, which promotes students' learning motivation for astronomical knowledge and allows them to become immersed in the simulation of the universe. This study mainly explored the following issues: (1) differences in astronomy learning performance between students in the upper grades of elementary school exposed to teaching with the AR-based Cosmos Planet Go App and those exposed to traditional classroom narrative teaching; (2) differences in learning motivation among students in the upper grades of elementary school between those using the AR-based Cosmos Planet Go App and those using traditional classroom narrative learning; (3) differences in flow experiences associated with astronomy curricula between students in the upper grades of elementary school who used the AR-based Cosmos Planet Go App and those who used the traditional classroom narrative learning; and (4) the analysis of creative perception with the AR-based Cosmos Planet Go App among students in the upper grades of elementary school.

2. Literature review

2.1. Creative situated learning and application

There have been many good research results on the introduction of situated learning theory, and they are applied most widely in the field of natural sciences. Annotation and comparison through field observation may help students understand the scientific phenomena described in abstract text (Tan et al., 2012). Huang et al. (2013) cited the life experience of convenience store shopping to guide elementary school students to learn addition and subtraction in mathematics. Hwang et al. (2011) constructed a Mindtool based on concept maps to assist learners in a ubiquitous situated learning environment. In the application of Chinese learning, some researchers have combined the difficult and incomprehensible collections of Chinese poems with the situation to deepen the learners' understanding of the meaning of their words and sentences (Shih et al., 2012). These studies have repeatedly shown that combining knowledge that is not easy for students to understand in the classroom with real simulated learning situations can achieve good learning results.

It is very important to support creativity to construct a learning environment that simulates the real situation through interaction. Many scholars have found that learning in a simulated real situation, co-creation and cooperation of learning climate and learners' ideas have been valued and discussed. These are necessary parts to support creativity in the process (Beghetto & Kaufman, 2014). Chen (2007) explored the influence of creative thinking teaching mode on adult English learning, which promoted adult learners to obtain autonomous learning and gain ability to cultivate problem-solving skills and communication and coordination skills through creative learning activities that simulate real situations. Hua (2020) explored the design thinking of museum creative learning resources research and development. It simulated the real situation of the digital learning resources through the actual participation and observation of researchers, and the questionnaire survey of learning resource users. The introduction of design thinking during the development of learning resources requires timely

interaction and discussion with participants, which help to transform abstract ideas into concrete content and help to generate ideas quickly.

2.2. Astronomy curriculum and augmented reality

Astronomy has been currently taught in elementary, junior high and high schools in most countries (Fleck & Simon, 2013). The boundless universe has always been the goal that people are striving to explore since ancient times, whose mystery also arouses the curiosity of most students, driving them to further explore the mystery of astronomy (Wu et al., 2015). Observing the operation of celestial bodies and astronomical phenomena is the teaching focus of astronomy curriculums in elementary and junior high schools (Plummer, 2014). However, if students want to observe specific celestial bodies, they have to match the time when the celestial bodies appear and the positions of observing the celestial bodies have also to be correct, which are quite difficult for students taking astronomy in school during the day. As a result, teachers need to take into account the limitations of time and space in teaching.

Many studies have applied AR in teaching. The study specifically designed an AR Chinese character learning game for young learners, and explored its impact on learners' cognitive engagement in classroom learning. The research results showed that students' cognitive engagement has increased significantly in AR-assisted learning activities (Wen, 2020). Chen (2020) combined AR and media with the real learning environment, and provided students with the scaffolding for constructing situated learning, so as to reduce the cognitive load of learners. Experimental results show that compared with traditional video teaching, AR multimedia video teaching method significantly improves students' learning effectiveness and intrinsic motivation, and improves students' satisfaction with English learning. Yousef (2021) explored the use of AR in the lower grades of elementary school students to enhance their creative thinking and increase the possibility of promotion in informal geometric training. The results found that the two research groups had significant differences in motivation and creative thinking skills. AR's presentation of learning content can enhance students' learning and motivate students' willingness to learn.

2.3. ARCS motivation model

The ARCS Motivation Model is proposed by Keller (1987) and explained how to design curriculums with the ARCS Motivation Model. In the element of Attention, it points out that it is necessary to change the previous teaching methods in order to grab students' interest, stimulate their inquiry into problems and arouse their curiosity; in the element of Relevance, it points out that the content of the curriculum needs to satisfy students' personal goals, meet their needs and increase their recognition of the curriculum by combining their previous learning or daily life experience; the element of Confidence illustrates the importance of setting success criteria so that students can realize that they can successfully complete the curriculum with their own efforts; and the element of Satisfaction explains the necessity of providing students with the opportunity to show their skills, teachers' giving students oral praise or substantive rewards after students' successful completing the learning goals set by the teachers, teachers and maintaining a fair reward mechanism. At present, many studies (Deublein et al., 2018; Turel & Sanal; 2018; Wu, 2018) used to combine ARCS teaching strategies with different courses to help students learn, and most study results also showed that applying the ARCS Motivation Model in teaching seemed to improve students' learning motivation and learning effectiveness. Therefore, this study took astronomy as the study subject to explore whether the use of AR-assisted learning system by the fifth and sixth graders will affect their learning motivation and learning effectiveness.

2.4. Flow experience

Flow hereby refers to a state of being, namely a process of one's being absorbed in an event or activity. When one enters the flow state, he/she will not feel bored with the event or activity in which he/she engages, but instead will be completely devoted to that event or activity (Csikszentmihalyi, 1975). Studies on flow were quite diverse, and many scholars (Chen et al., 2018; Yang & Quadir, 2018) also discussed people's behaviors in using new technology, games, shopping, learning, social networking sites and other activities. At the educational level, when a student has a higher flow state, he/she will become more efficient when learning; this is because one feels that learning is joyful and contented in his/her mind (Chang et al., 2017), which means that the more flow experienced one has, the more active one's learning behaviour will be (Hong et al., 2019). Nevertheless, it is really challenging for current education to enable students to achieve their flow states in learning (Ibáñez et al.,

2014). There were studies (Hong et al., 2019; Kao et al., 2019) showing that learning with multimedia-related textbooks can effectively improve students' flow state.

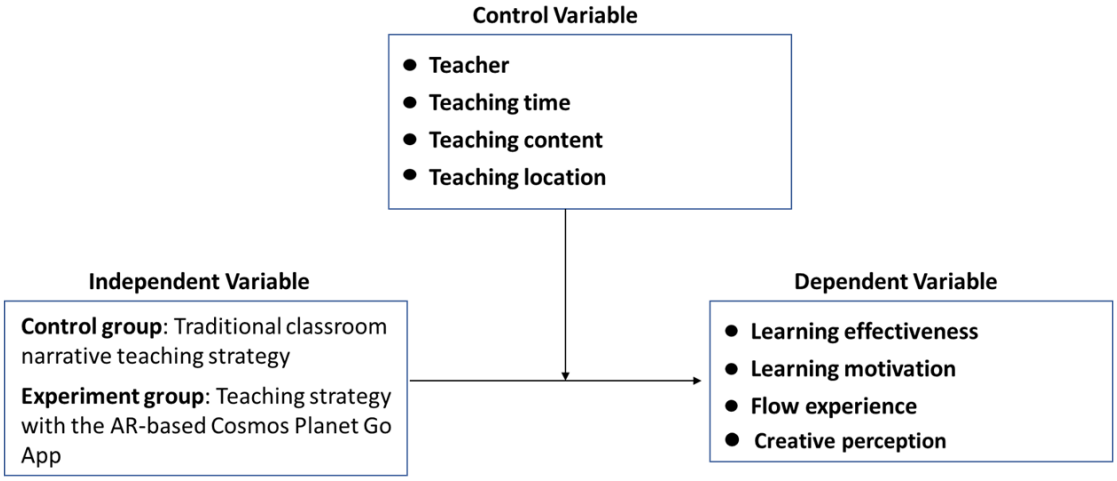
3. Research methodology

3.1. Experiment design

The research framework was shown in Figure 1. This study adopts the quasi-experimental method to conduct teaching experiments on astronomical knowledge in an elementary school of central Taiwan. The independent variable referred to the difference in the teaching strategies used in the experiment. The experimental group used the teaching strategy of AR-based Cosmos Planet Go App, while the control group used the traditional classroom narrative teaching strategy. The dependent variables were learning effectiveness, learning motivation, flow experience, and creative perception, and it discussed the differences in learning effectiveness, learning motivation, flow experience, and creative perception under different teaching strategies. In the control variables, teachers, teaching time, teaching content and teaching location of the two groups are the same. The operation definitions of learning motivation, learning effectiveness, flow experience and creative perception are described. Learning motivation refers to arousing the motivation of students to learn, and continuing to carry out learning activities, so that students' learning activities tend to the learning goals set by the teacher (Keller, 2010). Learning effectiveness refers to the changes in students' academic performance after participating in learning activities (Chen & Lin, 2016a). Flow experience refers to a state in which a person is completely immersed in a certain activity, ignoring the existence of other affairs. This kind of experience itself brings great joy and it is a subjective psychological feeling (Pearce et al., 2005). Creative perception uses the SCALE creative learning framework proposed by Richardson and Mishra (2018) to measure creative learning perception. The SCALE includes the classroom environment, classroom learning climate, and learner participation. Based on the above experimental design, the research hypotheses are described as follows:

- The learning effectiveness of the teaching strategy with the AR-based Cosmos Planet Go App is better than that of the traditional classroom narrative teaching strategy.
- The learning motivation of the teaching strategy with the AR-based Cosmos Planet Go App is higher than that of the traditional classroom narrative teaching strategy.
- The flow experience of the teaching strategy with the AR-based Cosmos Planet Go App is higher than that of the traditional classroom narrative teaching strategy.
- The creative perception of the teaching strategy with the AR-Based Cosmos Planet Go App is better than that of traditional classroom narrative teaching strategy.

Figure 1. The diagram of research framework



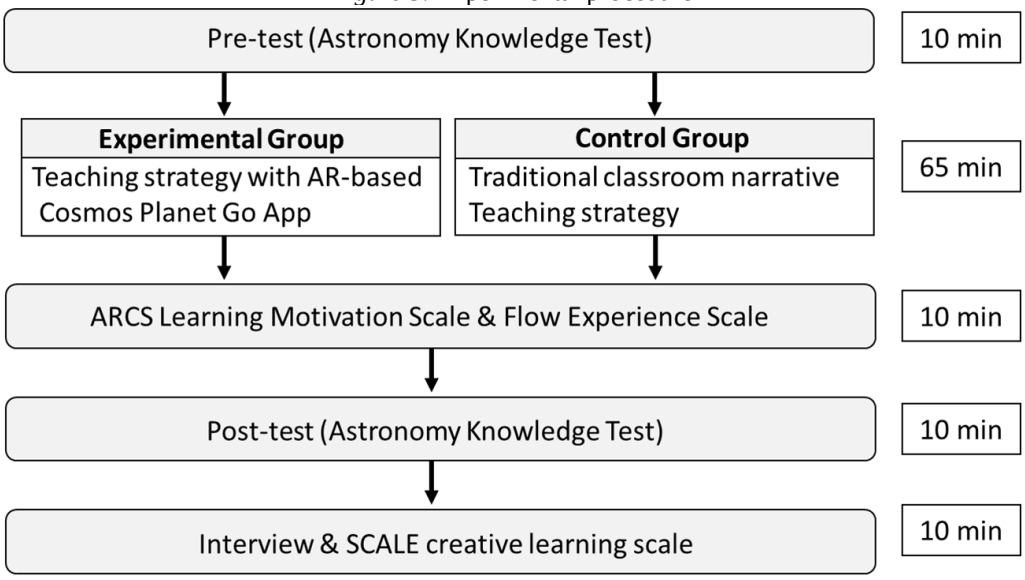
The participating students are students of the fifth and sixth grades because grades fifth and sixth of Taiwan elementary school have learned about the universe and planets. The experimental group and the control group are composed of students from one class in the fifth grade and one class in the sixth grade respectively. The Experimental group had 40 students, including 10 males and 10 females of fifth grade and 10 males and 10 females of sixth grade. The control group has 40 students, including 9 males and 10 females of fifth grade and 9 males and 12 females of sixth grade. The teaching course is the science course of elementary schools. The learning contents include the moon motion and rotation, the relationship between the moon orbit and the sun, the planets of the solar system and the distance of interstellar groups. The left side of Figure 2 simulates the situation

of planetary motion in the universe, and the right side of Figure 2 shows the positions of stars in the sky at night. The prior knowledge on astronomical knowledge possessed by the two groups of students is the consistent, and there is no significant difference.

Figure 2. Learning contents of the planets and stars

The experimental procedure is shown in Figure 3. The pre-test questions and post-test questions prepared were reviewed by three experts, and the questions with unclear or ambiguous meanings were deleted, and the exam questions were determined after item analysis and discrimination analysis. Item analysis can improve the quality of test questions, thereby improving the reliability and validity of the test. The greater the item-difficulty index, the easier the question is. Discrimination analysis is an assessment of the question proportion of papers that distinguish high-ability and low-ability. The higher the discrimination index is, the higher the consistency between the response of the subjects and the total score is (Hogan, 2007; Yu, 2011). After statistical analysis, the pre-test questions and post-test questions were of moderate difficulty and high discrimination. The teacher of the control group taught astronomical knowledge in an oral manner, and guided students through the pictures in the textbook and study sheets to strengthen students' exploration of learning astronomical knowledge, the connection of related knowledge concepts. Teachers gave timely feedbacks on incorrect concepts and rethought the study sheets, so as to achieve learning progress. The learning activities of astronomy curriculums for the students of control group and experimental group in this study are shown in Table 1.

Figure 3. Experimental procedure



All participating students were asked to fill in the IMMS (Instructional Materials Motivational Scale) and the flow experience test scale. They subsequently attended the astronomy achievement post-test. Lastly, the students in the experimental group were interviewed after class. The learning situation of the control group and the experimental group is shown in Figure 4. The IMMS scale developed by Keller (2010) has 4 dimensions included Attention, Relevance, Confidence and Satisfaction. The Cronbach's alphas of the 4 dimensions were 0.878, 0.861, 0.848 and 0.929 respectively. As for the Flow Experience Scale developed by Pearce et al. (2005) and was used to measure the students' overall flow experience. The Cronbach's alpha was 0.885, which was acceptable in reliability. A total of 20 students from the experimental group were randomly selected for after-class interviews. Each student spent an average of 8-10 minutes in the interview and filled in the SCALE

questionnaire. The interview content included the operation convenience of the AR-based Cosmos Planet Go App, learning experience and perception to creative learning environment.

Figure 4. Learning in the control and experimental groups

(a) (b)

Table 1. ARCS Motivation model curriculum design

ARCS Factor	Teaching strategy	
	Control group	Experimental group
Attention	1. Wonderful films and pictures, together with briefing teaching to grab students' senses. 2. From the content of teaching, ask questions to make students curious, and then trigger them to think. 3. The teaching briefing is presented with excellent videos and pictures, which is different from the previous teaching methods.	1. Combine AR with the simulated planets' rotation and revolution to arouse students' curiosity. 2. Let the students practice with the multiple-choice questions. After they answer the questions, the questions answered with wrong answers will be displayed on the tablet computer screen for them to understand which aspects of the content are still unclear to them, by which they are aroused to think. 3. The planets' appearances are presented in the form of AR together with dynamic visual presentation, which is distinguished from the previous teaching methods.
Relevance	1. Explain clearly the teaching objectives and learning priorities of astronomy curriculum. 2. Allow the students to ask questions. Explain the questions according to the questions raised by the students. Teachers may also ask the students questions. 3. The teaching content provides familiar examples to help students understand.	1. Clearly present the features and learning priorities of each planet 2. After the students answer the exercise questions, they are immediately shown with their answer results and all the correct answers on the tablet computer screen. 3. Each planet's features is described by the things that students are exposed to in their daily lives, helping them to understand the content more quickly.
Confidence	1. Define fair scoring items and criteria and establish objective rules and incentives. 2. The in-class worksheets are designed as the level of difficulty at which students can accomplish. 3. In the process of completing the in-class worksheets, students are provided with appropriate practice opportunities and feedback to help them complete the in-class worksheets.	1. The exercise questions have fair scoring items and criteria. If the student's answers are all correct, he/she will be given a verbal encouragement. 2. The exercise questions are designed to be a level of difficulty at which the students are capable of completing, and the App is easy to operate such that students can get started quickly. 3. Provide the students with appropriate practice opportunities and feedback to help them complete the exercise questions.

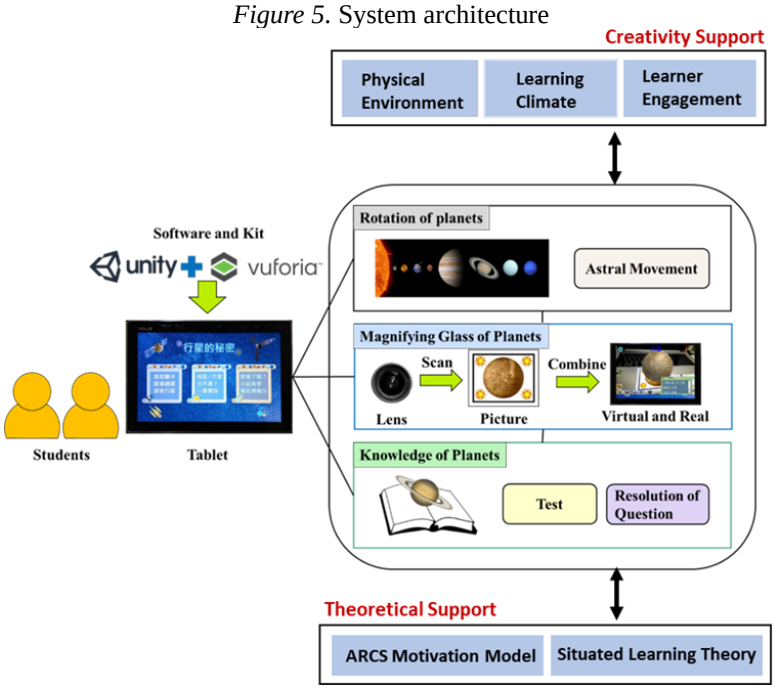
Satisfaction	1. Students complete the in-class worksheets according to the content of the curriculum. 2. Teachers provide students with useful information, feedback and timely reward. 3. Ensure fair scoring mechanism and rewarding criteria.	1. Students complete the exercise questions in the tablet computer according to what they have learned in the curriculum. 2. The App provides students with useful information, helpful feedback and timely rewards students. 3. Exercise questions have a fair scoring mechanism and rewarding criteria.

3.2. Design of AR-based Cosmos Planet Go App

Creative situated augmented reality learning refers to constructing a learning environment that simulates the real situation and supports creativity through AR technology. The creative situated augmented reality learning proposed is the integration of situated learning theory (Lave & Wenger, 1990), ARCS motivation theory (Keller, 1987) and creative learning environment model (Richardson & Mishra, 2018) as shown in Figure 5. The creative learning environment model includes physical environment, learning climate and learner engagement. The physical environment means that the learning environment should be open, allowing students in each group to collaborate and discuss with each other, and have a variety of rich learning content and resources available for students to read at any time (Peterson & Harrison, 2005; Warner & Myers, 2010). The learning climate refers to an open learning climate where students can freely discuss new ideas and trust each other. Creativity tends to flourish when there is an opportunity for exploration and learning and when innovation is valued (Kozbelt et al., 2010). Learner engagement refers to the learning tasks that students actually participate in. All members of a learning environment are seen as co-learners and co-teachers, emphasizing the importance of the learning process (Jeffrey & Craft, 2004). Learners can focus on questions and answer after thinking. These can be achieved by stimulating students' intrinsic motivation (Peterson & Harrison, 2005) or by training students to use various media to express ideas so that they can understand what the learner is actually doing.

The AR-based Cosmos Planet Go App implemented was written in Unity 3D and C# in the system interface and functions. The AR functions used Vuforia to develop engine, and 3D model materials of celestial bodies are obtained from the 3D Warehouse. Students use the system on a tablet to learn. The system architecture is shown in Figure 5.

Figure 5. System architecture



The learning App has three main functions: Planets' Revolution and Rotation, Magnifying Glass of Planets and Planets' Knowledge. In the function of Planets' Revolution and Rotation, it allows students to understand the planets' rotation, revolution and distance from the sun, where students can magnify, shrink and move by gesture touch so that they can clearly observe the planets' rotational directions and positional order. The system also

simulates the planet's rotation and revolution patterns, where students can click on the star icons on lower left corner for more detailed explanations. In the function of Magnifying Glass of Planets, it enables students to observe the appearance of planets carefully. As long as they pick up the tablet and scan the pictures of planets through the lens, the planets' 3D models will appear on the screen of the tablet and are integrated into the real environment. Students can magnify, shrink and rotate the planets through the gesture touch to find the features of each planet. They can also slide the planet information bar to learn about the planets. As in the function of Planets' Knowledge, it can help students to review. After completing the curriculum, students can answer the questions in tablet computers through gesture touch. Whether they answer correctly or incorrectly, the system will give students corresponding answers. Students can clarify the parts they do not understand so as to fulfil the review purpose. Each function on the system screen is shown in Figure 6.

Figure 6. System screen

In the construction of the creative situated augmented reality learning environment, students can not only use tablets to learn astronomy knowledge and discuss with peers but also use the intelligent computers built in the classroom to check planetary data at any time and experience the situational simulation operation of large-screen touch planets. During the learning process, students in different groups are allowed to discuss with each other, and each group is encouraged to put forward different ideas. Through the answers of different groups, the truth of astronomical knowledge can be verified, allowing students to explore knowledge and clarify the misunderstanding in their own learning concepts. When students put forward different points of view to express their feedback on the operation of the planet, the teacher will also help guide students to reflect, and then deepen the correct concept and increase the interest in participation. Through such a creative situated augmented reality learning environment learning process, students can express different creative thinking and experience planetary simulation at any time to achieve the positive effect of physical environment, learning climate, and learner engagement.

4. Experimental results

4.1. The influence of learning effectiveness

The results of independent samples *t*-test as shown in Table 2, there were no significant differences in the pre-test results between different groups, where $t(73.060) = -1.439$, $p = .155$. The students in the control group and the experimental group had the same understanding and awareness of astronomical knowledge before they participated in the learning activities. As shown in Table 3, the results of independent samples t-test presented that there were significant differences in the post-test results between different groups, where $t(62.073) = -2.367$, $p = .021$, indicating that the experimental group students had better learning effectiveness than the control group students did.

Table 2. t-test for different groups in their pre-test scores

	Mean (*SD*)		*t*	*p*
	Control group (*N* = 40)	Experimental group (*N* = 40)		
Pre-test	52.400 (33.675)	62.050 (25.806)	-1.439	.155

Table 3. t-test for different groups in their post-test scores

	Mean (*SD*)		*t*	*p*
	Control group (*N* = 40)	Experimental group (*N* = 40)		
Pre-test	82.000 (22.151)	91.550 (12.677)	-2.367	.021*

Note. *$p < .05$.

4.2. The influence of learning motivation

Independent samples *t*-test was applied to students' learning motivation in all 4 dimensions as shown as Table 4. In the dimension of Attention, there were significant differences between different groups, where $t(67.491) = -2.419$, $p = .018$; significant differences were found in the dimension of Relevance, where $t(78) = -2.411$, $p = .018$; significant differences were found in the dimension of Confidence, where $t(63.563) = -2.849$, $p = .006$; significant differences were also found in the dimension of Satisfaction, where $t(63.132) = -2.456$, $p = .017$; there were significant differences in ARCS, where $t(65.184) = -2.759$, $p = .008$. The results showed that the "AR-based Cosmos Planet Go App" can effectively grab students' motivation, inspire their inquiry into problems and arouse their curiosity, rendering them outperform the students learning by classroom narration teaching. In the classroom, students actively used the AR-based Cosmos Planet Go App, and they were particularly interested in the generated situation of the planetary motion connection in the universe. Students discussed the distance, temperature and weight of planets with each other. As they saw the simulated real AR-based planetary motion in the universe, students asked whether the planets may collide, and whether there may be planets that do not move according to the orbit. All these learning interactions showed that the teaching strategy of creative situated AR-based learning attracted students to concentrate on learning astronomical knowledge.

Table 4. t-test of different groups in the 4 dimensions of ARCS

	Mean (*SD*)		*t*	*p*
	Control group ($N = 40$)	Experimental group ($N = 40$)		
Attention	4.060 (0.994)	4.515 (0.655)	-2.419	.018*
Relevance	4.060 (1.039)	4.530 (0.663)	-2.411	.018*
Confidence	3.925 (1.097)	4.500 (0.653)	-2.849	.006**
Satisfaction	4.130 (1.088)	4.620 (0.640)	-2.456	.017*
ARCS	4.050 (0.961)	4.543 (0.597)	-2.759	.008**

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

4.3. The influence of flow experience

The analysis results of independent samples *t*-test as shown in Table 5, there were significant differences in the flow experience between different groups, where $t(72.598) = -13.912$, $p < .001$, indicating that after using the "AR-based Cosmos Planet Go App" to learn during the course, students were more immersed in their learning than those learning with the classroom narrative teaching way.

Table 5. t-test of different groups in their flow experience

	Mean (*SD*)		Mean (*SD*)	*p*
	Control group ($N = 40$)	Experimental group ($N = 40$)		
Flow	2.432 (0.450)	4.075 (0.596)	-13.912	< .001***

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

4.4. Interview analysis

This section discussed the interview analysis for participating the experiment process. Students (S05 and S12) said that they thought it was incredible to see the 3D planetary motion on the tablet, which can stimulate their motivation for learning and more creative thinking, and they could discuss with their classmates about the different characteristics and colors of each planet. Student (S03) said that in the motion of the planetary orbit, they can see Taiwan during the rotation by themselves. They unconsciously wanted to draw the method of finding the Polaris by themselves and provide it to classmates as a reference. Students (S07 and S17) said that this APP made it easier for them to learn planetary knowledge, and also made them actively want to participate in the learning activities of the course, and they wanted to recommend this APP to their good friends to learn together. Students (S09, S14 and S20) said that the system was very easy to use, especially the use of 3D animation to show the simulated motion of the real planet, which gave them more confidence in learning the knowledge of the planet, and also enabled them to discuss knowledge that was not understood with their classmates and review it again through the APP.

The SCALE developed by Richardson and Mishra (2018) was employed to measure students' perception of creative learning environment as shown in Table 6. In the dimension of physical environment, the item of PE1 received 85% of highly agree support from students. In the dimension of learning climate, the item of LC3 won

90% support from students. In the dimension of learner engagement, the items of LE4 and LE3 were supported by 95% and 90% for highly agree. According to the results, the most important students' perception about creative learning environment was to demonstrate an interest in or enthusiasm for the activity beyond being on task. Students' use of multiple perspectives/viewpoints/ways of knowing or various modes of investigation/problem solving is secondary importance. From the dimension of scale, learner engagement is the most important.

Table 6. Analysis of learners' perception for creative learning environments

Question		Mean	*SD*	Percentage of each question in 4-point Likert scale (%)			
				Disagree	Minimal Agree	Moderate Agree	Highly Agree
Physical Environment:							
PE1:	A variety of resources are available and accessible to students.	3.85	0.37	0%	0%	15%	85%
PE2:	Examples of student work appear in the space.	2.55	0.60	5%	35%	60%	0%
PE3:	A variety of work stations or areas are available to students.	3.15	0.75	0%	20%	45%	35%
PE4:	The furniture allows for multiple arrangements and configurations.	3.10	0.79	0%	25%	40%	35%
Learning Climate:							
LC1:	Students are involved in discussions among themselves, with or without the teacher, that deepen their understanding.	3.85	0.37	0%	0%	15%	85%
LC2	The students are caring, respectful, and value differences.	3.60	0.68	0%	10%	20%	70%
LC3:	The teacher is a facilitator, co-learner, explorer, or inquirer with students.	3.90	0.31	0%	0%	10%	90%
LC4:	Mistakes, risk-taking, and novel ideas are valued or encouraged.	3.20	0.52	0%	5%	70%	25%
Learner Engagement:							
LE1:	Students are involved in tasks that are open-ended and/or involve choice.	3.50	0.61	0%	5%	40%	55%
LE2:	Students are involved in activities that may include inquiry, project based learning, or interdisciplinary tasks.	3.60	0.50	0%	0%	40%	60%
LE3:	Students use multiple perspectives/viewpoints/ ways of knowing or various modes of investigation/problem solving.	3.90	0.31	0%	0%	10%	90%
LE4:	Students demonstrate interest in or enthusiasm for the activity beyond being "on task."	3.95	0.22	0%	0%	5%	95%
LE5:	Students spend time	3.30	0.73	0%	15%	40%	45%

	developing ideas for deeper understanding and/or reflecting on their learning.						
LE6:	Students work at their own pace and/or time is used flexibly.	3.85	0.37	0%	0%	35%	65%

5. Discussion

Many current studies have shown that the use of ARCS teaching strategy design can help students improve their learning effectiveness (Hung et al., 2013; Turel & Sanal, 2018; Wu, 2018). This study again verified that technology-assisted learning designed with ARCS teaching strategy can stimulate students' learning. According to the research results, teaching using AR-based Cosmos Planet Go App can improve students' learning motivation and learning effectiveness more than the traditional classroom narrative teaching method. Some past literature on digital technology-assisted learning has considered that digital technology-assisted learning environments support training of students' creative learning (Jahnke & Liebscher, 2020; Yeh et al., 2019). Digital technology-assisted learning focusing on the teaching scene can provide the guidance of teaching activities for students' creative learning and create a learning environment conducive to cultivating students' creative expression. This study proposed using situated augmented reality learning environment to support creative learning. Support for creative learning was constructed through an interactive learning environment that simulates real situations. Students discussed what they have learned and expressed their ideas creatively by simulating the real situation of how the stars operate in the universe.

This study proposed a creative situated augmented reality learning model to solve the students' problem in learning the abstract astronomy knowledge and focus on the key points of learning by simulating planetary operations through situated contexts. In addition to discussing learning effectiveness, learning motivation, and learning immersion, a creative situated augmented reality learning environment is the most important aspect of students' learning experience. This study found that the aspect of learning participation was the most important, which was not considered by many related literatures on the support of digital technology-assisted learning environments in the past. This study advocated the support of digital technology-assisted learning environments in terms of creative environment construction and implementation, the interactive E-learning system could focus on the modules that provide students with diverse perspectives to solve learning tasks, as well as the ARCS-based augmented reality curriculum design to promote students' participations in learning activities.

6. Conclusion

Augmented Reality can effectively enable students to observe and understand the course content. Teaching materials designed via ARCS teaching strategies can help students improve their learning effectiveness. This study proposed creative situated augmented reality learning environment combines situated learning theory, the characteristics of augmented reality with the ARCS motivation theory and creative learning environment model to develop the AR-based Cosmos Planet Go App. Students can understand the characteristics and features of each planet through this simulated universe motion situation. According to the analysis results, it can be seen that student learning with an AR-based Cosmos Planet Go App outperformed the ones learning only with the classroom narrative teaching method as the system can effectively assist students in their learning. The practical teacher in elementary schools can teach astronomy knowledges using the AR-based Cosmos Planet Go App to improve the learning effectiveness, learning motivation and enhance flow experience. Students can understand the characteristics and features of each planet through this simulated universe motion situation. At the academic contribution, this study proposed the creative situated learning with AR to solve the teaching in astronomy curriculums that is abstract and difficult to be understood. It constructed learning situations through AR so that learners can learn abstract concepts. It also enabled scholars to have more applications and discussions on the research topics of creative situated learning with AR.

Although more studies have discussed how digital technology-assisted learning environments support the construction and evaluation of students' creative learning, most of them have no clear principles on how to integrate students' creative learning into digital technology-assisted learning environments. In this study, we found that the dimension of learning engagement, a mode in which students could use multiple perspectives to solve the problems is the most important, and followed by students' showing interest or enthusiasm for learning

activities, instead of just performing tasks, as secondary. The results can be as a reference by future scholars with a priority to focus on supporting learner engagement when conducting research on digital technology-assisted learning environments supporting creative situations. Many scholars have successively discussed the design of teaching courses with the ARCS model. However, the most past literature took English as learning theme, and they used it to improve learning effectiveness by enhancing learning motivation. It is not easy to learn and understand astronomy knowledge without guidance by situation. This study used the AR to present the situation of planetary motion in the universe, so that students could understand astronomy knowledge clearly. Therefore, this study created the AR-based Cosmos Planet Go App support creativity and allowed students to observe the planetary motion that was usually difficult to observe at close range in real life. It can help teaching application of astronomy knowledge, and it may also enable practice teachers to have a variety of teaching methods.

7. Limitations and future research

It was suggested that the extension and analysis of different learning styles in the situated learning with AR can be added into the follow-up of this research, as well as what kind of learning interactivity should be possessed by learners of different learning styles in the situation constructed by AR, and differences in their influence on learning effectiveness, learning motivation and flow experience. In the experiment of this study, the 3D effect can only be effectively presented in a well-lit classroom because the simulated planet motion animation of augmented reality is presented on a tablet, If the classroom is lack of light, the effect of the augmented reality presentation may be affected. The AR-based Cosmos Planet Go App developed can only be used on tablet. Because of the different sizes of smartphone screens, it cannot effectively present virtual reality 3D animations that simulate real situations. Due to limited resources, the experiment time of this study was short, and the experiment time can be extended for future follow-up studies.

Acknowledgements

The authors wish to thank the Ministry of Science and Technology of the Republic of China for financially supporting this research under Contract No. MOST108-2511-H-005-001-MY3, and MOST105-2511-S-005-001-MY3.

References

Bressler, D. M., & Bodzin, A. M. (2013). A Mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *Journal of Computer Assisted Learning, 29*(6), 505-517.

Beghetto, R. A., & Kaufman, J. C. (2014). Classroom contexts for creativity. *High Ability Studies, 25*(1), 53–69

Chang, C. C., Liang, C., Chou, P. N., & Lin, G. Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game-based learning? Perspective from multimedia and media richness. *Computers in Human Behavior, 71*, 218-227.

Chen, H.-L. (2007). The Effect of creative teaching on adult english learners based on the ATDE model. *Bulletin of Diwan College of Management, 2*(1), 57-70.

Chen, C. C., & Lin, P. H. (2016a). Development and evaluation of a context-aware ubiquitous learning environment for astronomy education. *Interactive Learning Environments, 24*(3), 644-661.

Chen, C. H. (2020). AR videos as scaffolding to foster students' learning achievements and motivation in EFL learning. *British Journal of Educational Technology, 51*(3), 657-672.

Chen, H.-R., & Lin, Y.-S. (2016b). An Examination of digital game-based situated learning applied to Chinese language poetry education. *Technology, Pedagogy and Education, 25*(2), 171-186.

Chen, Y. M., Hsu, T. H., & Lu, Y. J. (2018). Impact of flow on mobile shopping intention. *Journal of Retailing and Consumer Services, 41*, 281-287.

Chin, K. Y., Lee, K. F., & Chen, Y. L. (2018a). Using an interactive ubiquitous learning system to enhance authentic learning experiences in a cultural heritage course. *Interactive Learning Environments, 26*(4), 444-459.

Chin, K. Y., Wang, C. S., & Chen, Y. L. (2018b). Effects of an augmented reality-based mobile system on students' learning achievements and motivation for a liberal arts course. *Interactive Learning Environments, 27*(7), 927-941. https://doi.org/10.1080/10494820.2018.1504308

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.

Deublein, A., Pfeifer, A., Merbach, K., Bruckner, K., Mengelkamp, C., & Lugrin, B. (2018). Scaffolding of motivation in learning using a social robot. *Computers & Education, 125*, 182-190.

Fleck, S., & Simon, G. (2013). An Augmented reality environment for astronomy learning in elementary grades: An Exploratory study. In *Proceedings of the 25th Conference on l'Interaction Homme-Machine* (pp. 14). ACM.

Hogan, T. P. (2007). *Educational assessment: a practical introduction*. John Wiley.

Hong, J. C., Tsai, C. R., Hsiao, H. S., Chen, P. H., Chu, K. C., Gu, J., & Sitthiworachart, J. (2019). The Effect of the "Prediction-observation-quiz-explanation" inquiry-based e-learning model on flow experience in green energy learning. *Computers & Education, 133*, 127-138.

Hua, C.-W. (2020). Design thinking and developing museum creative learning resources: The Mobile Museum of the National Museum of History. *Journal of Museum & Culture, 20*, 83-112

Hsu, T. C. (2017). Effects of gender and different augmented reality learning systems on English vocabulary learning of elementary school students. *Universal Access in the Information Society, 18*(2), 315-325.

Hwang, G.-J., Shi, Y.-R., & Chu, H.-C. (2011). A Concept map approach to developing collaborative Mindtools for context-aware ubiquitous learning. *British Journal of Educational Technology, 42*(5), 778-789.

Hwang, W.-Y., Zhao, L., Shadiev, R., Lin, L.-K., Shih, T. K., & Chen, H.-R. (2020). Exploring the effects of ubiquitous geometry learning in real situations. *Educational Technology Research and Development, 68*, 1121-1147.

Huang, S.-H., Wu, T.-T., Chen, H.-R., Yang, P.-C., Huang, Y.-M. (2013). Development of a context-aware game for conducting ubiquitous mathematics learning activities. *International Journal of Mobile Learning and Organisation, 7*(3/4), 239-252.

Hung, I. C., Chao, K. J., Lee, L., & Chen, N. S. (2013). Designing a robot teaching assistant for enhancing and sustaining learning motivation. *Interactive learning environments, 21*(2), 156-171.

Ibáñez, M. B., Di Serio, Á., Villarán, D., & Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education, 71*, 1-13.

Jahnke, I., & Liebscher, J. (2020). Three types of integrated course designs for using mobile technologies to support creativity in higher education. *Computers & Education, 146*, 103782. https://doi.org/10.1016/j.compedu.2019.103782

Jeffrey, B., & Craft, A. (2004). Teaching creatively and teaching for creativity: Distinctions and relationships. *Educational Studies, 30*(1), 77–87.

Kao, G. Y. M., Chiang, X. Z., & Foulsham, T. (2019). Reading behavior and the effect of embedded selfies in role-playing picture e-books: An Eye-tracking investigation. *Computers & Education, 136*, 99-112.

Keller, J. M. (1987). Strategies for stimulating the motivation to learn. *Performance & Instruction, 26*(8), 1-7.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. Springer.

Kozbelt, A., Beghetto, A. B., & Runco, A. M. (2010). Theories of creativity. In J. C. Kaufman & R. J. Sternberg (Eds), *The Cambridge Handbook of Creativity* (pp. 447-459). Cambridge University Press.

Lau, K.-W. (2011). The Importance of creative learning environment in art & design education. *Design Journal of Shih Chien University, 5*, 195-219.

Lave, J., & Wenger. E. (1990). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Li, K., & Keller, J. M. (2018). Use of the ARCS model in education: A Literature review. *Computers & Education, 122*, 54-62.

Pearce, J. M., Ainley, M., & Howard, S. (2005). The EBB and flow of online learning. *Computers in human behavior, 21*(5), 745-771.

Peterson, R. E., & Harrison III, H. L. (2005). The Created environment: An Assessment tool for technology education teachers: creativity doesn't just happen by chance; the prepared environment nourishes it. *The Technology Teacher, 64*(6), 7–11.

Plummer, J. D. (2014). Spatial thinking as the dimension of progress in an astronomy learning progression. *Studies in Science Education, 50*(1), 1-45.

Richardson, C., & Mishra, P. (2018). Learning environments that support student creativity: Developing the SCALE. *Thinking Skills and Creativity, 27*, 45-54.

Shih, W.-C., Tseng, S.-S., Yang, C.-C., Lin, C.-Y., & Liang, T. (2012). A Folksonomy-based guidance mechanism for context-aware ubiquitous learning: A Case study of Chinese scenic poetry appreciation activities. *Educational Technology & Society*, *15*(1), 90-101.

Stolaki, A, & Economides, A. A. (2018). The Creativity challenge game: An Educational intervention for creativity enhancement with the integration of Information and Communication Technologies (ICTs). *Computers & Education*, *123*, 195-211.

Tan, T.-H., Lin, M.-S., Chu, Y.-L., & Liu, T.-Y. (2012). Educational affordances of a ubiquitous learning environment in a natural science course. *Educational Technology & Society*, *15*(2), 206–219.

Tseng, H. W., Tang, Y., & Morris, B. (2016). Evaluation of iTunes university courses through instructional design strategies and m-learning framework. *Educational Technology & Society*, *19*(1), 199-210.

Turel, Y. K., & Sanal, S. O. (2018). The Effects of an ARCS based e-book on student's achievement, motivation and anxiety. *Computers & Education*, *127*, 130-140.

Warner, S. A., & Myers, K. L. (2010). The Creative classroom: The Role of space and place toward facilitating creativity. *The Technology Teacher*, *69*(4), 28-34.

Wen, Y. (2020). Augmented reality enhanced cognitive engagement: Designing classroom-based collaborative learning activities for young language learners. *Educational Technology Research and Development*, 1-18.

Wu, S. J., Chu, H. C., & Yang, K. H. (2015). An Web quest-based context-aware u-learning system to improve students' problem solving and communication abilities in astronomy inquiry activities. In *2015 IIAI 4th International Congress on Advanced Applied Informatics* (pp. 319-322). IEEE.

Wu, T. T. (2018). Improving the effectiveness of English vocabulary review by integrating ARCS with mobile game-based learning. *Journal of Computer Assisted Learning*, *34*(3), 315-323.

Yang, J. C., & Quadir, B. (2018). Individual differences in an English learning achievement system: Gaming flow experience, gender differences and learning motivation. *Technology, Pedagogy and Education*, *27*(3), 351-366.

Yeh, Y.-C., Rega, E. M., & Chen, S.-Y. (2019). Enhancing creativity through aesthetics-integrated computer-based training: The effectiveness of a FACE approach and exploration of moderators. *Computers & Education*, 139, 48-64.

Yousef, A. M. F. (2021). Augmented reality assisted learning achievement, motivation, and creativity for children of low-grade in primary school. *Journal of Computer Assisted Learning*, *37*(4), 966-977. https://doi.org/10.1111/jcal.12536

Yu, M.-N. (2011). *Educational test and assessment*. Psychological Publisher.

Zhang, J., Sung, Y. T., Hou, H. T., & Chang, K. E. (2014). The Development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction. *Computers & Education*, *73*, 178-188.

Appendix

Pretest Questions

1. What kind of planets are the stars on the constellation chart?
2. What are the main causes of craters on the surface of the Moon?
3. What is the order of the eight planets from nearest to farthest from the Sun?
4. If you see a rainbow in the East after a rain, which direction is the Sun located?
5. In the region where the Tropic of Cancer passes, which direction does the noon pole shadow face in winter?
6. Which planet is the largest in size?
7. Which planet has the obvious ring?
8. Which planet is covered in a thick cloud of sulfuric acid?
9. Does the feature of the moon going around the Earth mean that the side of the moon facing the Earth remains facing us?
10. What is the order of planetary rotation speed from slowest to fastest?

Posttest Questions

1. When the constellation is in motion, will the distance between the stars remain the same?
2. Observe craters on the surface of the moon. Is it true that the surface is damaged and uneven due to the heat of the moon's combustion as it orbits close to the sun?
3. What is the order of the eight planets from the sun from farthest to nearer?
4. In the area where the Tropic of Cancer passes, the sun rises from the northeast in the morning in the summer, and from which position will the sun set in the afternoon?
5. Is it true that planets usually revolve in the same direction as the star they orbit, and that the hotter the star's surface, the brighter it gets?
6. Which planet would float if placed in water?
7. Which planet is famous for its beautiful star rings, and is it the first planet to be discovered with star rings?
8. How long does it take for the moon to make a circle around the earth?
9. What is the order of the planet's rotation speed from fast to slow?
10. Which planet is characterized by the greenhouse effect?

Interview question

1. For learning with the AR-Based Cosmos Planet Go App, what is the help for astronomy learning?
2. For learning with the AR-Based Cosmos Planet Go App, what is the degree of improvement in learning motivation for participating in astronomy courses?
3. For learning with the AR-Based Cosmos Planet Go App, what is the degree of change in the learning immersion of participating in astronomy courses?
4. When learning with the AR-Based Cosmos Planet Go App, what is the ease of operation for using the system?

Effectiveness of Remote-Control Cars and Authentic Learning in Strengthening Creative Thinking and Problem-Solving Abilities

Ting-Ting Wu* **and Wei-Shan Liu**
Graduate School of Technological and Vocational Education, National Yunlin University of Science and Technology, Yunlin, Taiwan // ttwu@yuntech.edu.tw // d10943001@yuntech.edu.tw
*Corresponding author

ABSTRACT: Innovation in the design of curricula is widely discussed. Innovative curricula expose students to a diverse range of learning environments and prompt them to ask questions, stimulating their creativity and allowing them to develop a sense of initiative and to hone their problem-solving skills and ability to apply knowledge to practice. The introduction of new technology to the classroom has improved pedagogy and the information literacy of students. Because of these developments, this study expanded the integrative activity curriculum for second-grade elementary school students to an innovative curriculum involving a comprehensive set of activities related to remote-control cars and their use in the community. The students underwent a process of experiential learning in which they became familiar with the operation of remote-control cars. This study divided the students of two second-grade classes into an experimental group and a control group. The experimental group participated in innovative teaching activities as a part of authentic learning courses and were familiarized with the operations of remote-control cars in traffic in the community. The control group participated in innovative teaching activities as a part of the lesson plan for remote-control cars and were familiarized with the operation of the cars in traffic in the community. The creative thinking and problem-solving skills of the students in both groups significantly improved, and the students in the experimental group outperformed those in the control group. The students in both groups indicated that they were satisfied with the curriculum.

Keywords: Authentic learning, Curriculum design, Remote-control cars, Creative thinking, Problem solving

1. Introduction

Several innovative teaching methods have been introduced into schools' curricula, allowing for teachers and students to engage in various innovative learning approaches that strengthen their problem-solving and creative thinking skills (Hinkel, 2006; Williams, 2000). Authentic learning takes place in an environment in which students experience sensations through engaging situational learning (Herrington & Oliver, 2000). Learning in such an environment promotes creative thinking and higher levels of thinking in complex tasks, such as solving problems by analyzing, synthesizing, designing, manipulating, and evaluating information (Bath et al., 2004). This type of learning enables students to explore their own reactions and feelings rather than being constrained to a fixed curriculum (Maina, 2004; Zembylas, 2002). Creative thinking is an essential skill for the 21st century and its cultivation is therefore a primary pedagogical objective (Geisinger, 2016; Lee & Carpenter, 2015; Sternberg & Lubart, 1999). Critical thinking is expressed through words and ideas and applies not only to art and design but also to how we act, think, and relate to our environment (Mayer, 1989; Rhodes, 1987; Sternberg & Lubart, 1999). The most effective method for strengthening students' problem-solving skills is to let them investigate a problem by testing out various learning strategies (Goldschmidt & Smolkov, 2006) and to incorporate problem-solving techniques into educational activities (Seechaliao, 2017; Snyder & Snyder, 2008).

With the rise of cross-disciplinary education and Industry 4.0, from 2010 to 2020 in the United States, the number of mathematics teachers increased by 16%, the number of computer system analysts increased by 22%, the number of software designers increased by 32%, and the number of medical personnel increased by 36. The demand for biomedical engineers in two fields increased by 62%. These professionals must have basic cross-disciplinary skills (Vuong et al., 2019). Therefore, students should be acclimated to integrative technological thinking and logic throughout the learning process to improve their education (Roy et al., 2013). Cross-disciplinary technology education has been implemented in several curricular areas to transform experience into practice, encourage diversity and respect, and allow for students to reflect and expand on their learning experience (Levin & Nevo, 2009).

In authentic learning, students develop new ideas and approaches to solve problems. Brown et al. (2020) and Shadiev and Yang (2020) noted that educational technology, such as social networks, artificial intelligence, virtual robots, and wearable devices, can strengthen creative thinking and problem-solving skills. Authentic learning environments with a wide range of resources can stimulate students' creativity (Maina, 2004; Wu &

Wu, 2020). An authentic learning environment provides students with the space to use their imagination (Donovan et al., 1999) as well as opportunities to strengthen their creativity and problem-solving skills (Herrington & Oliver, 2000). These abilities are generally developed during later stages of education, such as the higher grades in elementary schools, high school, or university (Colbeck et al., 2000; Reinhold, 2006).

Yalçın and Erden (2021) described the effect of cross-disciplinary education activities in the design-thinking model on the creativity and problem-solving skills of preschool children. The use of small groups encouraged communication and interaction, and the children were able to apply what they learned in school to their lives. Çakır et al. (2021) noted that lessons on coding robots affected preschool children's problem-solving and creative thinking skills. Kaplancali and Demirkol (2017) provided programming courses to students aged ≥5 years and integrated mathematics into the curricular activities (Skemp, 1976). These studies demonstrate that many researchers have adopted a grounded approach to programming, allowing preschool and elementary schoolchildren to learn programming and coding through the educational methods in the curriculum and to develop an understanding of engineering technology. Although the aforementioned literature has reported that many courses for first and second graders have involved programming activities, few of them introduced the authentic learning approach. After the students acquire knowledge in class, they rarely have the opportunity to apply it to practice. In this study, a curriculum was designed to combine the activities in the school's curriculum with the opportunity to gain hands-on experience. Self-propelled vehicle assembly and programming activities were incorporated into the curriculum to increase students' engagement in learning and expand their range of knowledge and cognitive processes. Many studies have demonstrated the positive benefits of programming education for children. Otherwise, few studies have explored this in the context of second-grade elementary students or applied authentic curricula. Most studies have focused on programming education and computational thinking and have rarely explored creative thinking and problem solving.

Therefore, this study designed an innovative curriculum by adding lessons on remote-control cars to the integrative activity curriculum for second-grade students and investigated the efficacy of authentic learning in strengthening creative thinking and problem-solving skills. Authentic learning was implemented to provide students with a more experiential form of learning. The instructor for the experimental group followed an authentic curriculum in which students drew maps of their neighborhoods and then drew remote-control cars traveling through the map. The activity enabled the students to understand the concept of community living. In the programming lessons, the remote-control cars from the authentic curriculum were used to introduce the concept of driverless cars. The control group learned through the integration of remote-control cars into curriculum. The students' creative thinking, problem-solving abilities, and satisfaction with the course were then analyzed.

The research questions of this study were as follows:

- How would the students' creative thinking skills be affected in the experimental and control groups?
- How would the students' problem-solving skills be affected in the experimental and control groups?
- How satisfied would the groups be with the course?

2. Learning design

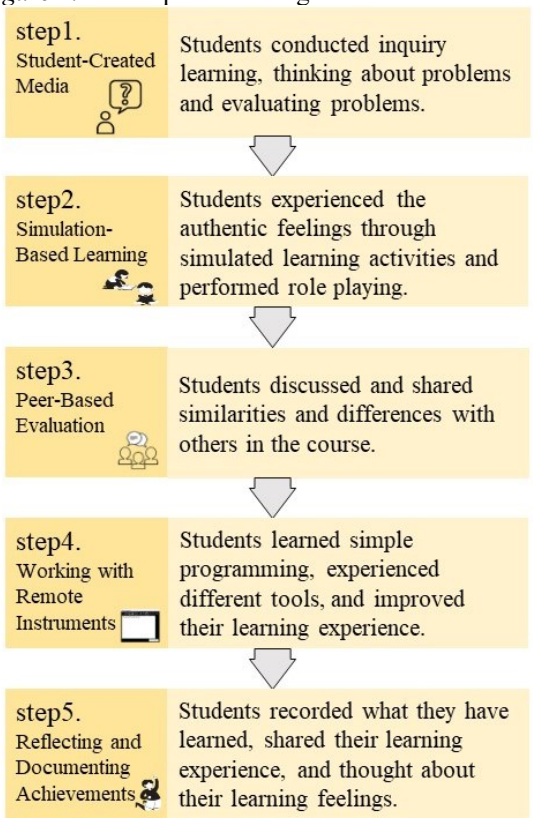
2.1. Considerations

Johnson et al. (2007) compiled examples of authentic learning in the 21st century. Researchers have identified 10 design elements of authentic learning that educators can apply to any subject area. We used five of these elements in the instructional design process, namely exploratory learning, simulated learning, peer evaluation, working with remote instruments, and reflection and documentation of achievements. Figure 1 presents the instructional model, with a break down of the learning process and framework.

The students engaged in exploratory learning, simulated learning, and peer evaluation. The students shared their experiences with their groups and the class throughout the course. In the first stage, the students participated in inquiry learning, asked questions, and reflected on problems to solve them through discussion with others using the knowledge they had gained. In the second stage, the students participated in simulated learning activities and engaged in role-play activities. Active participation in the course helped "develop valuable communication, collaboration, and leadership skills that help students succeed as professionals in their field of study" (Lombardi, 2008). In the third stage, the students identified similarities and differences. The students also participated in peer evaluation. Because each student has a unique perception of what is being taught, they can gain a deeper understanding of a subject through student–student discussion. In the fourth stage, the students learned simple

programming and used various technological tools (Lombardi, 2008). In the fifth stage, the students recorded what they had learned, shared their experiences, and reflected on their feelings during the process. The students also kept a record of their observations. The authentic learning tasks allowed for the students to reflect on how they learned (Blumenfeld et al., 1991; Lombardi, 2007; Newmann et al., 1996).

Figure 1. The steps of learning station rotation model



3. Research design

3.1. Curriculum design

This study explored the second-grade integrative activity curriculum of an elementary school in Taiwan. The integrative activity curriculum is intended to achieve the following: students learn topics related to self-knowledge, life management, social engagement, and environmentalism. The integrative activity curriculum consists of integrative classes on life, languages, health education, and arts and humanities. Students learn to respect life and multiculturalism, use and develop resources, and protect the environment. This experiment was conducted in two modules of the second-grade integrative activity course, namely "A Small Community in a Big World" and "Living Sphere in the Community." This curriculum consists of lessons on the community, living spheres in the community, and the connections between modes of transportation in that community. Remote-control cars were added to the original curriculum. The innovative curriculum was designed to enable students to familiarize themselves with remote-control cars and programming over the course of their education.

The experimental process was designed to enable the students to connect what they had learned to their daily lives. The experimental group received both modules of the integrative activity curriculum as well as some elements of authentic learning. The students drew a map of their community and then provided feedback. The map was later used for a simulated test drive of the remote-control car to deepen the students' understanding of the methods and logic of programming as applied to remote-control cars; the intention was that they would learn about how driverless cars detect objects, respond to verbal commands, and follow directions. The difference between the experimental group and the control group was that the students in the experimental group created an authentic link between the lesson and their own community. The concept of driverless cars was then introduced. A detailed description of the difference between the experimental and control groups is provided in Appendix 1.

3.2. Experimental process

3.2.1. Experimental group: Using driverless cars to integrate authentic learning

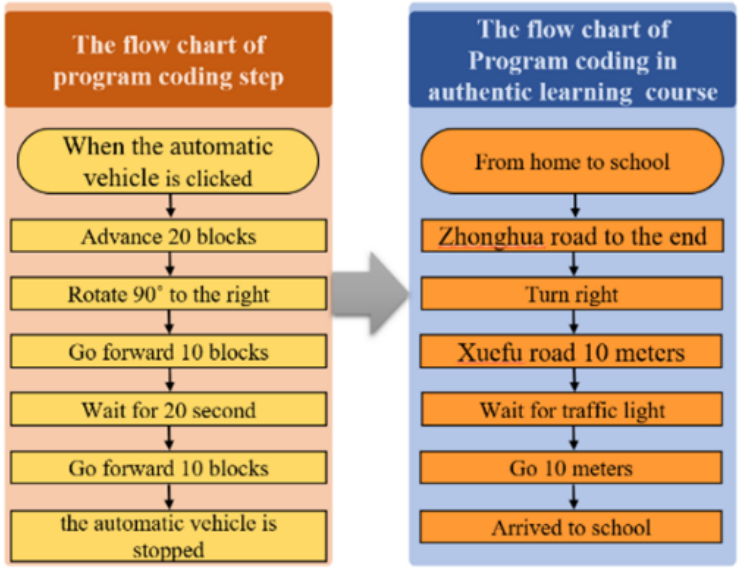
The experimental group completed a pretest for creative thinking and problem solving in the first week. The students were sorted into heterogeneous groups of four on the basis of their final exam scores the previous semester (24 students in six groups) for group discussion and sharing.

In the second week, the students began Modules 1 and 2 of the integrative activity curriculum, namely "A Small Community in a Big World" and "Living Sphere in the Community." The first module was intended to teach the students to understand their community and its connections other external communities. The second module was intended to teach the students to recognize the transportation tools used in their community; exploratory learning was encouraged through textbook activities. The students were asked to discuss the convenience of community life by exploring its connections with transportation and related problems. Then, the remote-control cars were introduced and integrated into the transportation and life activity.

During simulation learning in the third week, the experimental group engaged in simulation and role play by drawing community maps in authentic lessons. Simulation and role play allowed the students to actively participate in the curriculum games, learn how to operate the remote-control cars, and use examples from real life. Actual dialogues between drivers and passengers were integrated into the learning process. The teacher provided worksheets and textbook-based instruction. The worksheets were related to the lessons and questions from the textbook, and the students discussed and answered the questions together. During the sharing segments, the students were asked to write down their thoughts and feedback and share their feelings with their classmates. The students developed new perspectives by discussing their ideas with their classmates.

Weeks 4 and 5 involved peer evaluation and the remote-control car lesson. During the activities, the students worked with remote-control cars and were taught to operate the programs using standard programming procedures that incorporated real-life scenarios. Examples of these scenarios include "advance 20 blocks and turn left" and "continue to the traffic light, then turn left" (Figure 2). Students were encouraged to discuss problems and their strategies to operate the remote-control cars with each other. These authentic lessons enabled the students to perceive the innovativeness and convenience of driverless cars. The discussions allowed for the students to understand different viewpoints on topics and develop a deeper understanding of them.

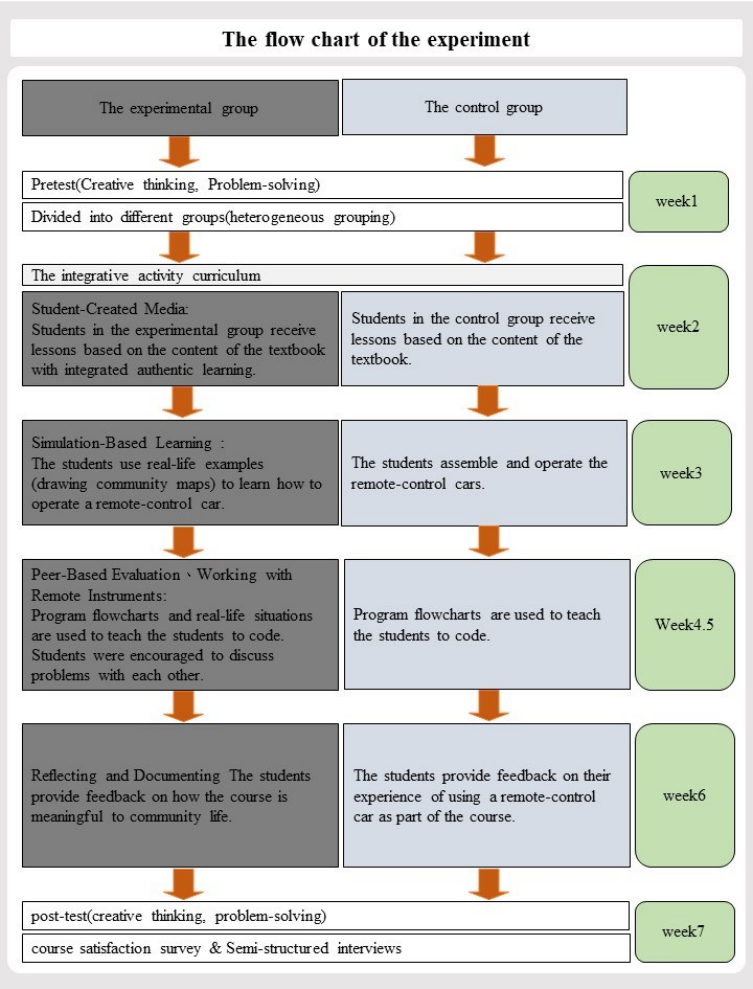
Figure 2. Real-life situation integrated into the standard program operation



In Week 6, the students provided feedback about the course after learning to control the remote-control cars remotely. The experimental group was asked to provide feedback on the lessons on community life, particularly in regards to the points they thought were meaningful and any connections they made. This gave the students an opportunity to reflect on their learning and to discuss their feelings about how the remote-control cars were incorporated into their lessons. The students also showed each other their worksheets and the remote-control cars they designed; this allowed the students to collectively reflect on their learning.

In Week 7, a post-test for creative thinking and problem solving was conducted, and a survey about the students' satisfaction with the course was distributed. The survey involved semi-structured interviews in which the students provided feedback on the community life lessons. Figure 3 presents a flowchart of the procedure, and the content of the semi-structured interviews is detailed in Appendix 2.

Figure 3. Experimental procedure



3.2.2. Control group: Using driverless cars to learn

The control group learned using the same curriculum as did the experimental group during Weeks 1, but the curriculum differed for Weeks 2–6. In Week 2, the control group began following Modules 1 and 2 of the integrative activity curriculum. Then, the remote-control car lessons were introduced.

In Week 3, the students assembled and operated the remote-control cars. The teacher provided textbook instruction and worksheets for the students to complete during class. The worksheets were related to the lessons and questions from the textbook, and the students discussed the answers with each other. During the student sharing segments, students were asked to write down their thoughts and feedback and to share their feelings with their classmates.

In Weeks 4 and 5, the lessons on the coding of the remote-control devices began. The classes in which the students learned how to operate the remote-control cars were taught using standard programming procedures. Students were able to discuss and express problems operating the remote-control cars while learning simple coding. The discussions helped the students to understand different viewpoints on the lesson and develop a deeper understanding of the topic. During the peer evaluation section and the teleoperation task, the students in the control group shared their experience and worked together to solve each other's problems.

In Week 6, the students provided feedback on their experience of using remote-control cars as part of this course. Week 7 was similar for the control and experimental groups, with the students completing a posttest for creative thinking and problem solving; a survey measuring the students' satisfaction with the course was also distributed.

3.3. Participants

Of the 48 participants (7–8 years old), 25 were boys and 23 were girls; the control and experimental groups each had 24 students. The experiment was approved by the school and the parents of the participants. All students who participated in this study were less than 18 years old. Thus, in accordance with ethical procedures, written consent was obtained from their parents, and all students expressed their willingness to participate. Appendix 3 presents the parental consent form.

Figure 4. Students completing the pretest and post-test

Figure 5. Students constructing the remote-control cars and completing the worksheets

Figure 6. Experimental group drawing maps of their neighborhood and sharing their ideas with classmates

The students had not previously participated in a similar course. The experimental group followed the innovative and authentic learning curricula, and the control group followed the innovative curriculum. Figure 4 depicts the students completing the pretest and posttest, and Figure 5 shows the students assembling the remote-control cars

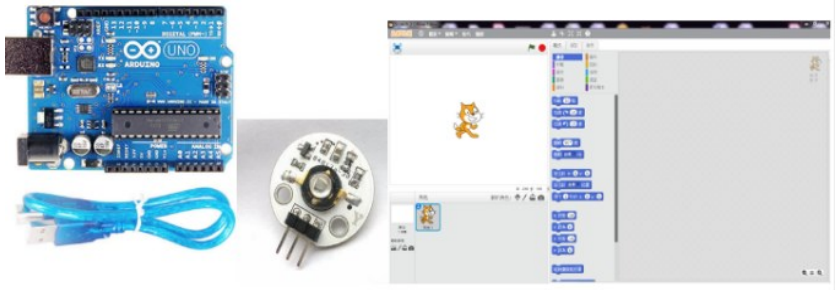
and filling out the worksheets. Figure 6 depicts those in the experimental group drawing maps of their neighborhood and working together during the integrative activity.

3.4. Devices used

Figure 7 displays the parts, remote control, and motor of the remote-control car. Figure 8 displays the Arduino Uno board, infrared transmitter, and operational interface of the programs to control the remote-control car. The building block car and its accessories were tools to help the students think logically, hone their creativity, develop motor skills, and cultivate habits of concentration. Once assembled, the block cars could be controlled remotely. To use the remote control, the students were required to use the control mode, which trains logical and mathematical thinking.

Figure 7. DIY Building Block (Remote-control) Car

Figure 8. Arduino Uno board, infrared Linker, and operational interface



4. Data collection and analysis

4.1. Torrance Tests of Creative Thinking

To evaluate creativity in the pretest and post-test, the study used the figural exercises of the Torrance Tests of Creative Thinking (TTCT), which comprises two tests. The pretest and post-test scores were divided into three domains: fluency, flexibility, and originality (Davis & Fichtenholtz, 2019). Fluency represents the ability to propose ideas in response to open-ended, oral, or nonverbal questions; flexibility represents the ability to adopt different methods in response to a problem, to consider different types of ideas, or to look at a situation from different angles; and originality represents statistical rarity or uniqueness and nonconformity. This study utilized Williams' Creativity Assessment Packet, with the creative-thinking activity used to establish criterion-related validity. The parameters were between .574 and .877. The internal correlation of the scores in Type A and Type B were between .597 and .812, all reaching a statistically significant level. Thus, the revised TTCT had favorable reliability and validity.

4.2. Problem-solving ability

Problem-solving ability was assessed through a revised version of a problem-solving test proposed in Bransford et al. (1986). The questions were related to five short stories and covered three domains: solutions, problem reasoning, and problem prevention. The participants answered the questions, and the reviewer analyzed the results for the three domains. The Cronbach's α for the test was 0.823. Solutions assessed participants' thinking ability in relation to proposing diverse and effective problem-solving concepts. Problem reasoning reflected

participants' thinking ability from various perspectives in their search for potential reasons for the problem. Problem prevention assessed participants' metacognitive abilities toward absorbing problem-solving experiences and formulating various methods to prevent problem occurrence.

4.3. Course satisfaction

Course satisfaction was measured through the questionnaire proposed by Alperin (1998) and Biner et al. (1997). The questions were scored on a 5-point Likert scale (*strongly agree* to *strongly disagree*). The questionnaire consisted of three parts, namely course content (four questions), self-identification (three questions), and teaching (three questions). The Cronbach's α for the questionnaire was 0.955, indicating considerably high internal consistency; thus, the questionnaire was reliable.

5. Research results

5.1. TTCT

Table 1 presents the two groups' independent-samples *t*-test scores for the creative thinking skills pretest. In the fluency domain, the mean score of the experimental group was 82.13 (standard deviation [*SD*] = 7.08), whereas that of the control group was 79.29 (*SD* = 9.47). The mean flexibility score was 81.17 (*SD* = 7.09) for the experimental group and 81.00 (*SD* = 6.91) for the control group. The mean originality score was 82.13 (*SD* = 7.81) for the experimental group and 81.33 (*SD* = 7.01) for the control group. The mean total score was 81.81 (*SD* = 6.83) for the experimental group and 80.50 (*SD* = 7.53) for the control group. For each item, $p > .05$ indicated no difference in creative thinking between the groups and that, therefore, the groups were homogenous.

Table 1. Independent samples *t*-test results for creative thinking pretest

Item	Group	Numbers	*M*	*SD*	*t*
Fluency	Experimental group	24	82.13	7.08	1.174
	Control group	24	79.29	9.47	
Flexibility	Experimental group	24	81.17	7.09	.082
	Control group	24	81.00	6.91	
Originality	Experimental group	24	82.13	7.81	.369
	Control group	24	81.33	7.01	
Total score	Experimental group	24	81.81	6.83	.629
	Control group	24	80.50	7.53	

According to the post-test independent-samples *t*-test results (Table 2), the mean fluency, flexibility, and originality scores of the experimental group were 86.41 (*SD* = 6.37), 86.08 (*SD* = 7.87), and 87.08 (*SD* = 7.67), respectively; their mean total score was 86.53 (*SD* = 7.09). The mean fluency, flexibility, and originality scores for the control group were 85.91 (*SD* = 8.53), 85.48 (*SD* = 8.81), and 86.54 (*SD* = 10.60), respectively; their mean total score was 85.98 (*SD* = 8.07). For each item, $p > .05$. A comparison of the results revealed no significant differences between groups; however, the mean score of the experimental group was higher than that of the control group.

Table 2. Independent samples *t*-test results for creative thinking post-test

Item	Group	Numbers	*M*	*SD*	*t*
Fluency	Experimental group	24	86.41	6.37	.970
	Control group	24	85.91	8.53	
Flexibility	Experimental group	24	86.08	7.87	.973
	Control group	24	85.48	8.81	
Originality	Experimental group	24	87.08	7.67	.901
	Control group	24	86.54	10.60	
Total score	Experimental group	24	86.53	7.09	.805
	Control group	24	85.98	8.07	

The results of a paired-samples *t* test indicated that the creative thinking ability of the experimental group significantly differed ($p < .001$) between the pretest and the post-test (Table 3). The *t*-test results for each part of the TTCT were as follows: fluency, −4.737 ($p = .001$); flexibility, −7.085 ($p = .001$); and originality, −7.536 ($p = .001$). The total TTCT score was −9.316 ($p = .000$).

Similarly, a paired-samples *t* test indicated that the creative thinking ability of the control group significantly differed ($p < .05$) in the pretest and post-test (Table 4). The *t*-test results were as follows: fluency, $t = -7.351$ ($p = .001$); flexibility, $t = -4.195$ ($p = .001$); originality, $t = -3.928$ ($p = .001$); and total TTCT score, $t = -6.374$ ($p = .001$).

Table 3. Paired samples *t*-test scores for creative thinking ability of experimental group

Item	Numbers	*M*	*SD*	*t*	*df*	*p*
Fluency	Pre-test(24)	82.13	3.71	-4.737	23	.001***
	Post-test(24)	86.42				
Flexibility	Pre-test(24)	81.17	3.40	-7.085	23	.001***
	Post-test(24)	86.08				
Originality	Pre-test(24)	82.13	3.22	-7.536	23	.001***
	Post-test(24)	87.08				
Total Score	Pre-test(24)	81.81	2.36	-9.316	23	.001***
	Post-test(24)	86.53				

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

Table 4. Paired samples *t*-test scores for creative thinking ability of control group

Item	Numbers	*M*	*SD*	*t*	*df*	*p*
Fluency	Pre-test(24)	79.29	4.89	-7.351	23	.001***
	Post-test(24)	85.91				
Flexibility	Pre-test(24)	81.00	6.03	-4.195	23	.001***
	Post-test(24)	85.48				
Originality	Pre-test(24)	81.33	7.59	-3.928	23	.001***
	Post-test(24)	86.54				
Total score	Pre-test(24)	80.50	4.21	-6.374	23	.001***
	Post-test(24)	85.98				

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

Although the two groups did not exhibit any significant differences in their post-test scores after the lessons with the remote-control cars, a comparison of the creative thinking pretest and post-test scores indicated significant differences in the three domains and in the total scores.

5.2. Problem-solving ability

Table 5 presents the pretest independent-samples *t*-test results for the problem-solving ability of the experimental and control groups. For solutions, the mean score was 23.50 ($SD = 4.20$) for the experimental group and 23.04 ($SD = 4.54$) for the control group. The mean problem reasoning score was 28.92 ($SD = 5.90$) for the experimental group and 28.96 ($SD = 5.59$) for the control group. The mean problem prevention score was 20.83 ($SD = 3.52$) for the experimental group and 21.08 ($SD = 3.61$) for the control group. The mean total score was 73.25 ($SD = 12.73$) for the experimental group and 73.08 ($SD = 13.07$) for the control group. For each item, $p > .05$ indicated that no significant differences were observed and therefore no heterogeneity was evident between the two groups.

Table 5. Independent samples *t*-test results for problem solving pretest

Item	Group	Numbers	*M*	*SD*	*t*
Solutions	Experimental group	24	23.50	4.20	.363
	Control group	24	23.04	4.54	
Problem reasoning	Experimental group	24	28.92	5.90	-.567
	Control group	24	28.96	5.59	
Problem prevention	Experimental group	24	20.83	3.52	.902
	Control group	24	21.08	3.61	
Total score	Experimental group	24	73.25	12.73	.964
	Control group	24	73.08	13.07	

Table 6 presents the post-test independent-samples *t*-test results for problem solving. The mean scores of the experimental group were 25.83 ($SD = 3.81$) for solutions, 31.96 ($SD = 4.98$) for problem reasoning, and 22.75 ($SD = 3.07$) for problem prevention. The mean scores of the control group were 24.25 ($SD = 4.19$) for solutions, 30.58 ($SD = 5.72$) for problem reasoning, and 22.75 (SD = 3.07) for problem prevention. The total mean score of

the experimental group was 80.54 ($SD = 10.89$), and that of the control group was 76.79 ($SD = 12.90$). For each item, $p > .05$ indicated no significant differences.

Table 6. Independent samples *t*-test results for problem solving post-test

Item	Group	Numbers	*M*	*SD*	*t*
Solutions	Experimental group	24	25.83	3.81	.178
	Control group	24	24.25	4.19	
Problem reasoning	Experimental group	24	31.96	4.98	.379
	Control group	24	30.58	5.72	
Problem prevention	Experimental group	24	22.75	3.07	.416
	Control group	24	21.96	3.59	
Total score	Experimental group	24	80.54	10.89	.282
	Control group	24	76.79	12.90	

A paired-samples *t* test indicated a significant improvement ($p < .001$) in the problem-solving ability of the experimental group (Table 7). The results (*t*) for problem solving were as follows: solutions, −5.02 ($p = .001$); problem reasoning, −6.50 ($p = .001$); and problem prevention, −4.86 ($p = .001$); the total score for problem solving was −6.84 ($p = .001$).

Table 7. Paired samples *t*-test results for problem-solving ability of experimental group

Item	Numbers	*M*	*SD*	*t*	*df*	*p*
Solutions	Pre-test(24)	23.50	2.277	-5.02	23	.001***
	Post-test(24)	25.83				
Problem reasoning	Pre-test(24)	28.92	2.293	-6.50	23	.001***
	Post-test(24)	31.96				
Problem prevention	Pre-test(24)	20.83	1.932	-4.86	23	.001***
	Post-test(24)	22.75				
Total problem-solving score	Pre-test(24)	73.25	5.221	-6.84	23	.001***
	Post-test(24)	80.54				

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

A paired-samples *t* test indicated a significant improvement ($p < .001$) in the problem-solving ability of the control group (Table 8) as well. The results (*t*) were as follows: problem solving, −6.06 ($p = .001$); problem reasoning, −4.23 ($p = .001$); and problem prevention, −3.84 ($p = .001$); the total score for problem solving was −5.50 ($p = .001$).

Table 8. Paired samples *t*-test results for problem-solving ability of control group

Item	Numbers	*M*	*SD*	*t*	*df*	*p*
Solutions	Pre-test(24)	23.04	0.977	-6.06	23	.001***
	Post-test(24)	24.25				
Problem reasoning	Pre-test(24)	28.96	1.884	-4.23	23	.001***
	Post-test(24)	30.58				
Problem prevention	Pre-test(24)	21.08	1.116	-3.84	23	.001***
	Post-test(24)	21.96				
Total problem-solving score	Pre-test(24)	73.08	3.303	-5.50	23	.001***
	Post-test(24)	76.79				

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

After the integrative activity modules, both the experimental group and the control group had significant improvements in problem-solving abilities.

5.3. Course satisfaction

Table 9 presents the independent-samples *t*-test results for course satisfaction of the experimental and control groups. The experimental group had a mean value of 4.24, whereas the control group had a mean value of 4.13. Under equal variance, the *t* value was nonsignificant ($t = 0.745$, $p = .461$), and *p* was greater than .05, indicating that the groups did not have significant differences in course satisfaction. The means in each area of satisfaction were higher than 4.

Table 9. Independent-samples *t*-test results for course satisfaction

Dimension	Group	N	*M*	*SD*	*t*	*p*
Course content	Experimental group	24	4.34	.619	1.889	.065
	Control group	24	3.98	.710		
Self-identification	Experimental group	24	4.17	.697	.080	.936
	Control group	24	4.15	.652		
Teacher teaching	Experimental group	24	4.15	.655	-.158	.875
	Control group	24	4.18	.800		
Overall satisfaction	Experimental group	24	4.24	.440	.745	.461
	Control group	24	4.13	.432		

6. Discussion

6.1. Improvements in creative thinking abilities

This study incorporated remote-control cars into integrated activity lessons. The students were instructed to use the cars on the maps of their neighborhoods they had drawn as a method of combining technology into the integrative activity curriculum. The activities combined programming with explorations of new environments, enabling students to learn from practice and to practice while learning (Dorst & Cross, 2001). The curriculum was designed to promote the creative thinking skills of second-grade students through peer discussions and integrative learning involving the assembly of remote-control cars, ideation, and programming (Hasanah & Surya, 2017). Creative thinking skills were assessed in terms of fluency, flexibility, and originality. In the post-test, total creative thinking scores significantly improved; therefore, the introduction of the remote-control cars strengthened the students' creative thinking skills. People with flexible thinking ability exercise diverse thinking processes and can make inferences. Therefore, this course diversified students' thinking.

As Syahrin et al. (2019) noted, technology courses using remote-control cars can improve creative thinking skills. The use of remote-control cars in the curriculum was a similar approach as that in other curricula incorporating elements of programming to facilitate the learning of advanced concepts and to strengthen students' creative problem-solving skills (Johnson et al., 1994; Rahmawati et al., 2019; Torrance et al., 1970).

According to the results, the interactive methods used by the experimental group, such as map drawing and group discussions, helped the students develop new, creative ideas (Webb et al., 2006). Davidson and O'Leary (1990) determined that programming classes in which students can practice elements of design improve creative thinking skills and learning motivation. The use of authentic materials (Rego et al., 2012) helps students to learn quickly in integrative activities involving remote-control cars or programming (Madden et al., 2013) and promotes active engagement, collaborative learning, and peer discussions, which enhance creative thinking and ideation. The process of assembling remote-control cars and engaging in discussion, exploration, and discovery with peers and teachers to find solutions enabled the students to have a unique experience using their creative thinking skills (Palanica et al., 2019; Taylor, 2016).

In the control group, the mean score for creative thinking skills increased considerably. This result demonstrates the effectiveness of lessons using remote-control cars to develop creative thinking skills. The innovative curriculum introduced several opportunities for the students to engage in creative thinking (Capraro & Nite, 2014), stimulated their creativity, and strengthened their problem-solving skills (Li & Yang, 2009). Both groups engaged in discussion, programming, exploration, and discovery while assembling the remote-control cars, which improved their performance in the creative thinking assessments; the results of the independent-samples *t*-tests indicated that the difference between groups was nonsignificant. The results also indicate the benefits of the techniques, viewpoints, and methods the students adopted to solve problems for their creative thinking abilities; such abilities are essential in science and engineering (Murcia et al., 2020).

The total post-test scores indicated significant improvements in both the experimental group and the control group; therefore, the programming activities, whether through map drawing or the use of the cars, were successful in developing the students' creative thinking. Although the post-test scores of the experimental and control groups did not reach statistical significance, the mean scores of each item in the experimental group were higher than those of the control group. In the comprehensive activity course of constructing driverless cars, students used authentic course learning. They also produced an authentic map. Regarding originality, they broadened their unique and creative perspectives and, in terms of fluency, exercised flexible and coherent thinking, formulating multiple feasible ideas. As for flexibility, they could make inferences and offered different,

unconventional views. These results were reflected in their mean scores. This finding is consistent with those of Shatunova et al. (2019) and demonstrates that the use of integrative learning in programming lessons strengthens the creative skills of second-grade students.

6.2. Improvements in problem-solving abilities

According to Vygotsky (1978), concepts must be integrated into cognitive structures in the appropriate social environment. He emphasized that social interaction is crucial in the learning of advanced thinking skills (Choi & Hannafin, 1995). This view is supported by the results of this study, which demonstrated that experiential learning and map drawing improved the higher-level thinking skills of the students in the experimental group (Greenstein, 2012). The students assisted each other to complete the map-drawing activity (Gillies & Haynes, 2011), systematically organized knowledge, and shared their discoveries and ideas for solving problems in discussions (Nickerson & Zenger, 2004). After the map drawing and remote-control car lessons, the students' problem-solving abilities improved. Although the control group was not taught coding through map drawing, the introduction of integrative coding activities in the integrative activity curriculum significantly improved their problem-solving abilities (Çiftci & Bildiren, 2020). By assembling and studying the remote-control cars, both groups improved their problem-solving skills through exploration, discovery, and discussion.

The groups exhibited significant improvements in their total scores for problem solving and in their scores for all three domains, namely solutions, problem reasoning, and problem prevention. Thus, the remote-control car lessons significantly affected the acquisition of basic knowledge and techniques and were effective in the domains of solutions, problem reasoning, and problem prevention (Johnson & Johnson, 1987; Johnson et al., 1984). The change in the problem prevention scores in the experimental group was more significant than that of the control group. The experimental group also had higher mean scores in all areas of the post-test. This demonstrates the effectiveness of introducing elements of authentic learning to problem-solving tasks (Popat & Starkey, 2019).

The activities strengthened the students' logical thinking, improved their judgment and reasoning, and enhanced their problem-solving skills considerably (Schunk et al., 1987). The results also indicated that creative thinking and problem-solving skills are inextricably linked (Siegle, 2017). The curriculum helped the students strengthen their creative thinking and problem-solving skills (Tuomi et al., 2018). The innovative design of the integrative activity curriculum enabled students to learn through discussion and interactions with their peers (Huang, 2019), which significantly improved their problem-solving skills.

6.3. Student satisfaction

Both groups reported high levels of satisfaction with the course, the process of self-identification, and the teaching. The groups did not significantly differ in terms of satisfaction. Therefore, programming tasks did not have a significant effect. The semi-structured interviews conducted at the end of the course revealed that students thoroughly enjoyed the course. Although some students found the course challenging, they were still enthusiastic about the activities. Appendix 2 documents some of the interview content.

7. Conclusions and future research

The results of the TTCT and problem-solving test indicated that the course improved the creative thinking and problem-solving abilities of both groups. Thus, teachers should encourage students to participate in a diverse range of activities to hone their creativity, increase their interest in a subject, and expand their knowledge; higher levels of engagement result in more effective learning. Students experienced different learning methods. Regardless of whether authentic course learning was applied, students were presented with the correlation between driverless cars and the course and understood this correlation in the context of the course, community, and life. Moreover, they learned simple programming to solve life problems. Future research can investigate correlations with other dimensions of creativity and problem solving, such as students' abilities to engage in trial and error, evaluate a problem comprehensively, and understand the contents of a lesson. A detailed description of the analysis of the results of the experimental and control groups is provided in Appendix 4.

Because of the time constraints, the lesson-course lasted only 7 weeks, In the future, if time allows, the experimental duration may be extended, which may reveal more substantial differences between the

experimental group and control group. Demonstrative inquiry is more likely to be effective if the lessons are held over a longer period of time in which students have more opportunities to practice. This study involved only 48 student participants, and the interviews produced only a small amount of data. In the future, if the number of students was increased, the amount of interview data would likely be larger. Future research can also investigate the relationship between integrative programming curricula and problem-solving and critical thinking skills. When delivering integrative courses, instructors should identify connections between creative thinking, problem-solving, and critical thinking skills. This type of research can elucidate the effects of different education practices on the development of creativity.

Acknowledgement

The research reported in this paper has been supported in part by the Ministry of Science and Technology (MOST) in Taiwan under the research project number MOST 110-2622-H-224-001 -CC2, MOST 108-2628-H-224 -001 -MY3, and MOST 110-2511-H-224 -003 -MY3.

References

Alperin, D. E. (1998). Factors related to student satisfaction with child welfare field placements. *Journal of Social Work Education*, *34*(1), 43-54.

Bath, D., Smith, C., Stein, S., & Swann, R. (2004). Beyond mapping and embedding graduate attributes: Bringing together quality assurance and action learning to create a validated and living curriculum. *Higher Education Research and Development, 23*(3), 313-328.

Biner, P. M., Welsh, K. D., Barone, N. M., Summers, M., & Dean, R. S. (1997). The impact of remote-site group size on student satisfaction and relative performance in interactive telecourses. *American Journal of Distance Education*, *11*(1), 23-33.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational psychologist, 26*(3-4), 369-398.

Bransford, J., Sherwood, R., Vye, N., & Rieser, J. (1986). Teaching thinking and problem solving: Research foundations. *American psychologist*, *41*(10), 1078-1089. https://doi.org/10.1037/0003-066X.41.10.1078

Brown, M., McCormack, M., Reeves, J., Brook, D. C., Grajek, S., Alexander, B., Bali, M., Bulger, S., Dark, S., Engelbert, N., Gannon, K., Gautheir, A., Gibson, D., Gibson, R., Lundin, B., Veletsianos, G., & Weber, N. (2020). *2020 Educause Horizon Report: Teaching and Learning Edition.* EDUCAUSE.

Çakır, R., Korkmaz, Ö., İdil, Ö., & Erdoğmuş, F. U. (2021). The Effect of robotic coding education on preschoolers' problem solving and creative thinking skills. *Thinking Skills and Creativity*, *40*, 100812. https://doi.org/10.1016/j.tsc.2021.100812

Capraro, M. M. & Nite, S. B. (2014). STEM integration in mathematics standards. *Middle Grades Research Journal, 9*(3),1-10.

Choi, J. I., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology Research and Development, 43*(2), 53-69.

Çiftci, S., & Bildiren, A. (2020). The Effect of coding courses on the cognitive abilities and problem-solving skills of preschool children. *Computer Science Education, 30*(1), 3-21.

Colbeck, C. L., Campbell, S. E., & Bjorklund, S. A. (2000). Grouping in the dark: What college students learn from group projects. *The Journal of Higher Education, 71*(1), 60-83.

Davidson, N., & O'Leary, P. W. (1990). How cooperative learning can enhance mastery teaching. *Educational Leadership, 47*, 30-33.

Davis, T. J., & Fichtenholtz, H. M. (2019). Thinking Outside of Functional Fixedness with the Aide of Mental Fatigue. *Creativity Research Journal*, *31*(2), 223-228.

Donovan, S. M., Bransford, J. B., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice*. National Research Council National Academy Press.

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem–solution. *Design Studies, 22*(5), 425-437.

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them? *Applied Measurement in Education, 29*(4), 245-249.

Gillies, R. M., & Haynes, M. (2011). Increasing explanatory behaviour, problem-solving, and reasoning within classes using cooperative group work. *Instructional Science, 39*(3), 349-366.

Goldschmidt, G., & Smolkov, M. (2006). Variances in the impact of visual stimuli on design problem solving performance. *Design Studies, 27*(5), 549-569.

Greenstein, L. M. (2012). *Assessing 21st century skills: A Guide to evaluating mastery and authentic learning*. Corwin Press.

Hasanah, M. A., & Surya, E. (2017). Differences in the abilities of creative thinking and problem solving of students in mathematics by using cooperative learning and learning of problem solving. *International Journal of Sciences: Basic and Applied Research (IJSBAR), 34*(01), 286-299.

Herrington, J., & Oliver, R. (2000). An Instructional design framework for authentic learning environments. *Educational Technology Research and Development, 48*(3), 23-48.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly, 40*(1), 109-131.

Huang, T. C. (2019). Seeing creativity in an augmented experiential learning environment. *Universal Access in the Information Society, 18*(2), 301-313.

Johnson, D. W., & Johnson, R. T. (1987). *Learning together and alone: Cooperative, competitive and individualistic learning*. Prentice-Hall.

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1994). *The New circles of learning: Cooperation in the classroom and school*. Association for Supervision and Curriculum Development.

Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, *19*(1), 15-29. https://doi.org/10.1007/s10648-006-9038-8

Johnson, D. W., Johnson, R. T., Holubec, E. J., & Roy, P. (1984). *Circles of learning: Cooperation in the classroom*. Association for Supervision and Curriculum Development.

Kaplancali, U. T., & Demirkol, Z. (2017). Teaching coding to children: A Methodology for kids 5+. *International Journal of Elementary Education*, *6*(4), 32-37.

Lee, S., & Carpenter, R. (2015). Creative thinking for 21st century composing practices: Creativity pedagogies across disciplines. *Across the Disciplines, 12*(4), 1-24.

Levin, T., & Nevo, Y. (2009). Exploring teachers' views on learning and teaching in the context of a trans-disciplinary curriculum. *Journal of Curriculum Studies, 41*(4), 439-465.

Li, M. H., & Yang, Y. (2009). Determinants of problem solving, social support seeking, and avoidance: A Path analytic model. *International Journal of Stress Management*, *16*(3), 155-176. https://doi.org/10.1037/a0016844

Lombardi, M. M. (2007). Authentic learning for the 21st century: An Overview. *Educause learning initiative, 1*(2007), 1-12.

Lombardi, M. M. (2008). Making the grade: The role of assessment in authentic learning. *EDUCAUSE Learning Initiative*, 1-16.

Madden, M. E., Baxter, M., Beauchamp, H., Bouchard, K., Habermas, D., Huff, M., Ladd, B., Pearon, J., & Plague, G. (2013). Rethinking STEM education: An Interdisciplinary STEAM curriculum. *Procedia Computer Science, 20*, 541-546.

Maina, F. W. (2004). Authentic learning: Perspectives from contemporary educators. *Journal of Authentic Learning, 1*(1), 1-8.

Mayer, R. E. (1989). Cognitive views of creativity: Creative teaching for creative learning. *Contemporary educational Psychology, 14*(3), 203-211.

Murcia, K., Pepper, C., Joubert, M., Cross, E., & Wilson, S. (2020). A Framework for identifying and developing children's creative thinking while coding with digital technologies. *Issues in Educational Research, 30*(4), 1395-1417.

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*(4), 280-312.

Nickerson, J. A., & Zenger, T. R. (2004). A Knowledge-based theory of the firm—The Problem-solving perspective. *Organization Science, 15*(6), 617-632.

Palanica, A., Lyons, A., Cooper, M., Lee, A., & Fossat, Y. (2019). A Comparison of nature and urban environments on creative thinking across different levels of reality. *Journal of Environmental Psychology, 63*, 44-51.

Popat, S., & Starkey, L. (2019). Learning to code or coding to learn? A Systematic review. *Computers & Education, 128*, 365-376.

Rahmawati, Y., Ridwan, A., & Hadinugrahaningsih, T. (2019). Developing critical and creative thinking skills through STEAM integration in chemistry learning. *Journal of Physics: Conference Series, 1156*(1), 012033. https://doi.org/10.1088/1742-6596/1156/1/012033

Rego, A., Sousa, F., Marques, C., & e Cunha, M. P. (2012). Authentic leadership promoting employees' psychological capital and creativity. *Journal of Business Research, 65*(3), 429-437. https://doi.org/10.1016/j.jbusres.2011.10.003

Reinhold, S. (2006). WikiTrails: Augmenting wiki structure for collaborative, interdisciplinary learning. In *Proceedings of the 2006 international symposium on Wikis* (pp. 47-58). https://doi.org/10.1145/1149453.1149467

Rhodes, M. (1987). An Analysis of creativity. In S. G. Isaksen (Ed.), *Frontiers of creativity research: Beyond the basics* (pp. 216–222). Bearly.

Roy, S. D., Mei, T., Zeng, W., & Li, S. (2013). Towards cross-domain learning for social video popularity prediction. *IEEE Transactions on multimedia, 15*(6), 1255-1267.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*(1), 54-61. http://dx.doi.org/10.1037/0022-0663.79.1.54

Seechaliao, T. (2017). Instructional strategies to support creativity and innovation in education. *Journal of Education and Learning, 6*(4), 201-208.

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

Shatunova, O., Anisimova, T., Sabirova, F., & Kalimullina, O. (2019). STEAM as an innovative educational technology. *Journal of Social Studies Education Research, 10*(2), 131-144.

Siegle, D. (2017). Technology: Encouraging creativity and problem solving through coding. *Gifted Child Today, 40*(2), 117-123.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*(1), 20-26.

Snyder, L. G., & Snyder, M. J. (2008). Teaching critical thinking and problem solving skills. *The Journal of Research in Business Education, 50*(2), 90-99.

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

Syahrin, A., Suwignyo, H., & Priyatni, E. T. (2019). Creative thinking patterns in student's scientific works. *Eurasian Journal of Educational Research, 19*(81), 21-36.

Taylor, P. C. (2016). Why is a STEAM curriculum perspective crucial to the 21st century? In *14th Annual conference of the Australian Council for Educational Research*, 89-93. https://research.acer.edu.au/research_conference/RC2016/9august/6

Torrance, E. P., Gowan, J. C., Wu, J. J., & Aliotti, N. C. (1970). Creative functioning of monolingual and bilingual children in Singapore. *Journal of Educational Psychology*, *61*(1), 72-75. https://doi.org/10.1037/h0028767

Tuomi, P., Multisilta, J., Saarikoski, P., & Suominen, J. (2018). Coding skills as a success factor for a society. *Education and Information Technologies, 23*(1), 419-434.

Vuong, Q. H., Ho, M. T., Vuong, T. T., La, V. P., Ho, M. T., Nghiem, K. C. P., Tran, B. X., Giang, H.-H., Giang, T.-V., Latkin, C., Nguyen, H.-K. T., Ho, C. S. H., & Ho, R. C. M. (2019). Artificial intelligence vs. natural stupidity: Evaluating AI readiness for the vietnamese medical information system. *Journal of Clinical Medicine, 8*(2), 168. https://doi.org/10.3390/jcm8020168

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

Webb, N. M., Kersting, N., & Nemer, K. M. (2006). Help seeking in cooperative learning groups. In *Help seeking in academic settings* (pp. 56-99). Routledge.

Williams, P. J. (2000). Design: The Only methodology of technology? *Journal of Technology Education, 11*(2), 48-60.

Wu, T. T., & Wu, Y. T. (2020). Applying project-based learning and SCAMPER teaching strategies in engineering education to explore the influence of creativity on cognition, personal motivation, and personality traits. *Thinking Skills and Creativity, 35*, 100631. https://doi.org/10.1016/j.tsc.2020.100631

Yalçın, V., & Erden, Ş. (2021). The Effect of STEM activities prepared according to the design thinking model on preschool children's creativity and problem-solving skills. *Thinking Skills and Creativity*, 100864. https://doi.org/10.1016/j.tsc.2021.100864

Zembylas, M. (2002). "Structures of feeling" in curriculum and teaching: Theorizing the emotional rules. *Educational theory, 52*(2), 187-208.

Appendix 1

Difference between the experimental group and control group in terms of the curriculum design, content, and process.

<table>
<tr><th></th><th>Experimental group</th><th>Control group</th></tr>
<tr><td>Curriculum design and tools</td><td>1. Remote-control cars
2. Learning coding (Scratch)
3. Authentic learning</td><td>1. Remote-control cars
2. Learning coding (Scratch)</td></tr>
<tr><td>Curriculum content</td><td>1. Authentic learning is integrated into the course through the use of remote-control cars.
2. The process of programming is integrated into real-life situations.</td><td>1. Using the remote-control cars to learn.
2. Program coding using a general program chart.</td></tr>
<tr><td>Curriculum schedule</td><td>
<table>
<tr><th>Week</th><th>Learning process</th></tr>
<tr><td>2</td><td>Students in the experimental group receive lessons based on the content of the textbook with integrated authentic learning.</td></tr>
<tr><td>3</td><td>Simulation-based learning: The students use real-life examples (drawing community maps) to learn how to operate a remote-control car.</td></tr>
<tr><td>4.5</td><td>Peer-based evaluation: Program flowcharts and real-life situations are used to teach the students to code. Students were encouraged to discuss problems with each other.</td></tr>
<tr><td>6</td><td>The students provide feedback on how the course is meaningful to community life.</td></tr>
</table>
</td><td>
<table>
<tr><th>Week</th><th>Learning process</th></tr>
<tr><td>2</td><td>Students in the control group receive lessons based on the content of the textbook.</td></tr>
<tr><td>3</td><td>The students assemble and operate the remote-control cars.</td></tr>
<tr><td>4.5</td><td>Program flowcharts are used to teach the students to code.</td></tr>
<tr><td>6</td><td>The students provide feedback on their experience of using a remote-control car as part of the course.</td></tr>
</table>
</td></tr>
</table>

The difference between the experimental and control groups:
The learning process for the experimental group integrated authentic learning and remote-control car programming to allow students to make connection between their daily life and the course content. The learning process for the control group centered on remote-control car programming, enabling students to learn programming.

Appendix 2

Semi-structured interviews

The recordings of the interviews were transcribed, and textual analysis was performed. For the first topic—"Personal learning experience: what do you think you have gained from the innovative classes?"— responses frequently focused on whether the content was engaging as well as on time management and course flow. The following are sample responses from three students:
E20-1: "I think it's interesting to explore and study mobile cars."
C10-1: "The time management was good, but some students weren't familiar with the software, so they always asked the teacher, and the classroom was a little noisy."
E18-1: "It took a lot of time to learn the material in this new course, so I think it was difficult."

The second interview topic was "Inter-team relationships: Do you think peer learning helps you learn effectively?" The responses predominantly referenced "cooperation between classmates," "group discussion," and "sometimes the discussion was too noisy." However, some students reported that discussion improved their learning experience. The following are representative responses:
E6-2: "Receiving help from my classmates improved my learning ability."
C5-2: "Discussion can bring more ideas into the course."
C9-2: "Sometimes, the discussion was so noisy that we could not think about the course content."

The third interview topic was "Experience with technology: Did you find it challenging to use the remote-control devices?" The feedback was focused on using the car, and numerous students commented that learning to use the remote control was easy or challenging and that the task was difficult. Example responses from four students are as follows:
E14-3: "I think it was easy to learn the program. However, I think it was a challenge for me to learn [how to use] the computer and remote control cars. Sometimes, I could not do it."
E8-3: "It was difficult for me. However, if it had been connected to an interesting real-life situation, it might have been easier to learn to control it."
C11-3: "It was alright because the teacher explained it in detail during class."
C22-3: "It was challenging."

Appendix 3

Ver. 2017.04.24, Human Research Ethics Committee of National Cheng Kung University

Consent Form for the Child's Parent or Legal Representative

Dear parents:
We are a research team from the Graduate School of Technological and Vocational Education at National Yunlin University of Science and Technology working under Dr. Ting-Ting Wu, the director of the research program. We would like to invite your child to assist us in understanding whether the integration of self-propelled vehicle activities into comprehensive activity courses can improve students' creative thinking and problem-solving abilities. Your child's participation would benefit other children and their parents. In addition to experiencing new learning methods, your child will learn to absorb knowledge and skills through interdisciplinary teaching, preparing him or her for future challenges.

The research goal:
The 21st century is an era of rapid change and development. To increase the competitiveness of Taiwan, the cultivation of talent has received increasing attention. Numerous countries have introduced educational concepts or reform measures to develop new indicators for children's learning. The remote-control car program involves hands-on problem-solving and exploration-oriented teaching, which can cultivate children's comprehensive abilities, including inquiry, critical thinking, creative thinking, and problem-solving. While solving problems and exploring solutions, the children will inevitably encounter mistakes and failure. At such times, they will reflect on the cause, correct their mistakes, try again, fail again, reflect again, and try again until they succeed. The remote-control car program can assist children in cultivating their patience, willpower, and responsibility. The program will be applied to lower grades to explore its effects on creative thinking, problem-solving, and critical thinking.
Procedure:
(1) Pretest and posttest: Before and after course, creative thinking and problem-solving tests will be conducted

to evaluate the effects of the program.
(2) Remote-control car program: The program allows students to participate in interdisciplinary activities in different courses, thus improving their creative thinking.
(3) Interviews: In one or two classes before the end of the course, a 20-min interview will be conducted with the children regarding their experience and take-aways from the course. For data accuracy, the interviews will be recorded. If you would not like the interview to be recorded or wish to end the interview at any point, please notify us. After the course, you (and those who withdraw from the study) will receive stationery as a token of our gratitude.
(4) Recording: We will request that review the verbatim transcript of the audio. We assume the responsibility of confidentiality. Your real name will not be used in the results, and we will deidentify the data to the best of our ability in resulting publications. However, your identity may be disclosed in unexpected circumstances; therefore, please carefully consider the terms of the interview. The recordings and verbatim transcripts will be stored on a hard disk or computer with a password in the laboratory of Modern Learning Technologies and Applications and will be deleted 5 years after the program (July 31, 2027); they will only be used in this study. If you are interested in the results of this study, a summary of the report can be provided.
(5) Your and your child's information will remain confidential
1. We will request that the students' teachers, who will serve as assistants in this study, distribute the consent form and study materials. The experimental activities will be conducted in both the course implemented by this study and school's normal course. The study course will be integrated into the school course regardless of your child's participation in the study. If you would not like to have your child participate, we will not include your child's learning and test results in our results. This course will not affect your child's academic performance or the teacher's perception thereof. This experimental test and your child's academic performance are separate and unrelated.
2. We will adopt an anonymized approach to the publication and results related to this study and will replace the children's real names with codes.

*** We will fulfill our responsibility to protect and respect your children.**
Participation in this study will not cause physical or psychological harm to the children. We will fully respect your decision should you wish to withdraw your child from the study. If you have any questions after the interview, please contact us. The course will be interdisciplinary and based on themes from the students' class. If a participant withdraws from the experiment, we will continue the study as planned, but the activities and test results of those who withdraw will be excluded from the study data. In such a case, we and the teachers will score your child separately from those remaining in the study. However, the scoring systems will be unrelated.

Human Research Ethics Committee of National Cheng Kung University

※ **Small gifts:**
During the teaching process, if the students are willing to share what they have learned from the course or exhibit strong learning performance, we will reward them with stationery as encouragement.
Please feel free to ask any questions regarding this form. If you agree to have your child participate, please complete the following section and provide your signature. Please do not feel pressured to participate in this study.
Signature of parent or legal representative:
Pretest and posttest: □Agree □Disagree
Remote-control car activities: □Agree □ Disagree
Interview recording: □ Agree □Disagree
Report of results: □Not required □Please mail a report after the study to the following address: _____
Signature: _____ Date: MM/DD/YY
Signature of the research team:

Ethical approval for this study was obtained from the Human Research Ethics Committee of National Cheng Kung University commissioned by National Yunlin University of Science and Technology. If you wish to discuss the rights and interests of the participants in this study or file a complaint, please contact the Committee. Tel: E-mail:

□This consent form is in duplicate and will be retained by both parties for their record.
Signature of program director/codirector/researcher: _____Date: MM/DD/YY

Appendix 4

Analysis of results for experimental and control groups:

1. The pretest and posttests scores of the experimental and control groups significantly differed, indicating that the use of the remote-control cars in the course improved the students' creative thinking and problem-solving abilities.
2. Although the comparison of the experimental and control groups in the creative thinking skills test is not significant, it can be found from the experimental activity that the application of remote-control cars can improve the creativity of students. The learning of creativity requires a long period of training and learning. Perhaps the experiment time can be extended to see more results of analysis.
3. A comparison of the average scores of the experimental and control groups for creative thinking revealed that the scores for fluency, flexibility, and originality were higher in the experimental group than in the control group. Although only a slight difference in scores was observed, the use of authentic learning in the course allowed the students to draw community maps, which, along with the course activities, increased their fluency, flexibility, and originality in creative thinking.
4. Although the comparison of the experimental and control groups in the problem solving test is not significant, it can be found that both groups need to carry out problem solving activities in the course. It can be found from the results that both groups are improved problem solving skills after taking the course.
5. A comparison of the average scores of the experimental and control groups for problem solving revealed that the scores for solutions, problem reasoning, and problem prevention were not significantly different after the course. However, through authentic learning, the students in the experimental group used their hand-drawn community maps to connect their learning with their environment. This real-life connection significantly increased the scores for solutions, problem reasoning, and problem prevention.
6. The authentic learning component ensured that the course incorporated real-life scenarios, allowing the students to apply their knowledge. Because the experiment lasted only 5 weeks, detecting significant improvements in creative thinking and problem solving is difficult. In the future, the experiment can be extended to observe differences more clearly.
7. In terms of course satisfaction, both the experimental group and the control group are quite satisfied with the course learning, and there is no significant difference between the two groups. It can be found that the two groups of students are satisfied with the course arrangement, learning progress and course content.