



Comparative Learning Performance and Mental Involvement in Collaborative Inquiry Learning: Three Modalities of Using Virtual Lever Manipulative

Cixiao Wang¹ · Yuying Ma² · Feng Wu³

Published online: 12 June 2020
© Springer Nature B.V. 2020

Abstract

In scientific inquiry learning, manipulatives have been widely utilized as learning resources. Studies have explored the advantages of virtual manipulative (VM) for conceptual understanding and knowledge construction in science education. However, research on the mental engagement and perception of students during collaborative learning under different modalities of using VM remains rare. In this study, we designed a virtual lever manipulative (VLM) and three modalities of using VLM in a primary science course: one VLM per student, one VLM per group, and one VLM per class. There were 80 fifth graders from three classes who participated in this quasi-experimental study. They were asked to complete a group worksheet during collaborative learning activities. Cognitive load, as well as flow experience, was investigated through a questionnaire survey after the learning activities. Task involvement was calculated by using the mental effort dimension of cognitive load and post-test scores. The findings indicate that class B (one VLM per group) gained the highest scores in group worksheet and the post-test 1 week later, followed by class C (one VLM per class) and class A (one VLM per student). Class B had the highest level of task involvement, as they had a shared screen among group members. In contrast, class A had a relatively low task involvement and spent more time checking consistency. Besides, both classes A and B had a higher sense of flow experience than class C. Class C experienced a traditional teaching method and less interaction with learning materials, leading to a lower level of flow experience and moderate task involvement.

Keywords Science education · Virtual manipulative · Mobile learning · Inquiry learning · Collaborative learning

Introduction

Hands-on activities using manipulatives have been widely used in science courses for teaching scientific concepts and phenomena (Koning and Tabbers 2011). With the

development of information technologies, virtual manipulatives (VMs) designed by simulating physical manipulatives (PMs) emerged (Moyer et al. 2002), such as the virtual labs that can help students conduct procedures of physics experiments (e.g., Zacharia and Olympiou 2011). As a digital learning tool, VM can work as an alternative way when physical resources are limited in classrooms (Bouck and Flanagan 2009). Studies have explored how VM works in science education and found that the augmentation of experimental phenomena, repeatable operations, and timely feedback of VM could benefit students' conceptual understanding and knowledge construction (e.g., Zacharia and Olympiou 2011; Olympiou and Zacharia 2012; Wang and Tseng 2018; Kapici et al. 2019). Meanwhile, as the availability of mobile devices in classroom settings increases, the applications of virtual manipulatives in science inquiry learning are becoming flexible and convenient (Min et al. 2016; Crompton et al. 2016). Studies have found that 1:1 student-mobile device ratio could enhance primary students' learning experience, because of the affordances of mobile technologies for supporting

✉ Feng Wu
wufeng@pku.edu.cn

Cixiao Wang
wangcixiao@yeah.net

Yuying Ma
664374468@qq.com

- ¹ School of Educational Technology, Beijing Normal University, Beijing, China
- ² Science Education and Research Group, Beijing Yuxiang Primary School Huilongguan School, Beijing, China
- ³ Department of Educational Technology, Graduate School of Education, Peking University, No.5 Yiheyuan Road Haidian District, Beijing 100871, People's Republic of China

learning in and outside of class (e.g., Looi et al. 2011; Wong and Looi 2011). However, in face-to-face collaborative inquiry learning, students in groups are usually asked to operate one set of physical manipulatives together. In this case, we might need to think about the ways of providing VM for collaborative inquiry learning.

During face-to-face collaboration, manipulatives have been proven to promote peer interaction and the construction of shared knowledge, as students can obtain and exchange information by operating manipulatives (Cáceres et al. 2018). The number of manipulatives in groups could make a difference in information exchanges. Meanwhile, the amount and type of support offered to students will affect the ability to carry out the task and thus will also affect their collaborative learning experience (Kirschner et al. 2018). Under the circumstances, when adopting VM based on mobile devices in collaborative inquiry learning, the student-mobile device ratio needs to be taken into consideration. However, former studies focus more on the comparison of using VM and PM in collaborative inquiry activities (e.g., Yuan et al. 2010; Zacharia and Olympiou 2011), or different combinations of VM and PM for inquiry learning (e.g., Kapici et al. 2019). Whether the student-mobile device ratio will influence students' learning experience and performance when engaging group inquiry work has not been addressed yet. Therefore, in this study, we developed a VM for students to explore the law of lever in groups and it can be run on web browsers. Three modalities of using VM for collaborative inquiry learning are proposed: one VM per student, one VM per group, and one VM per class. To compare students' learning performance and mental involvement under the three conditions, we conducted an educational quasi-experiment in a primary school.

Literature Review

Virtual Manipulatives in Science Education

Virtual manipulatives (VMs) are often modeled after physical manipulatives (PMs) and it can be accessed via the internet or computer software on portable devices (Moyer 2001). By adding interactive features and providing students with unlimited quantities to use with a problem, VM could be used as an alternative tool for teaching and learning scientific knowledge, especially when PM is limited in science classrooms (Bouck and Flanagan 2009). As mentioned by Olympiou and Zacharia (2012), the affordances of VM are as follows: observing phenomena that cannot be observed in real life, making or repeating accurate measurements or operations, and overcoming time-consuming procedures. Studies have explored the advantage of VM over PM on students' conceptual understanding and knowledge construction (e.g., Ha and Fang 2017; Zacharia and Olympiou 2011). Compared with PM,

VM could provide more efficient manipulation, repeatable experiments, and timely feedback, so that students could spend more time on conceptual understanding and conclusion discussion (Zacharia et al. 2008).

In traditional classrooms, the ways of adopting VM for inquiry learning may differ from what had been already tested in experiment settings (Cáceres et al. 2018). For example, in some situations, there is only a projection screen in a classroom to run a VM, where students just watch how their teacher manipulates learning objects in front of the classroom (Moyer 2001). In other situations, students are allocated to make operations on one VM in groups, which is similar to the way of adopting PM for group work (e.g., Zacharia et al. 2008; Ha and Fang 2017). Meanwhile, there is a global interest regarding the use of mobile technologies in science education (Crompton et al. 2016). The increasing availability of mobile devices in classrooms brings new ways of using VM in science inquiry learning (Min et al. 2016). It has been found that one mobile device per student could support seamless learning (Wong and Looi 2011). The ownership of mobile devices in a classroom enables students to explore freely and become more responsible for their own learning (Looi et al. 2011). In this case, the student-mobile device ratio seems to be an important factor for the usage of VM based on mobile devices in collaborative inquiry learning. However, studies related to comparisons of different ways of providing VM for groups remain rare, especially when VM runs on mobile devices.

Collaborative inquiry learning involves the distribution and interaction of information among members (Bell et al. 2010). Different modalities of using VM may affect the information exchange and task coordination in the process of cooperation. Studies have found that student-mobile device ratio could influence the interaction among group members when engaging in collaborative learning. For instance, Lin et al. (2012) compared two modalities of using tablets in collaboration on concept mapping: some groups work together with only one tablet (1:m); others work together with one tablet per student (1:1). The findings of the study showed that the 1:m condition resulted in lower perceptions toward tablet use and collaborative learning, with generated superior artifacts as all the notes were well discussed among the group members (Lin et al. 2012). In contrast, 1:1 condition demonstrated more consistency in group participation, which improved communication and interaction among group members (Lin et al. 2012). From this point of view, different modalities of using technologies in collaboration may influence students' learning performance and experience. Thus, there is a need to make a comparison among different ways of using VM for collaborative learning in science courses.

Mental Involvement During Collaborative Learning

The investigations of the instructional design of tasks and resources with tablets continue flourishing, which facilitates

students' learning experience (Mulet et al. 2019). Collaborative learning is one of the most frequent types of mobile learning activities in K-12 settings (Crompton et al. 2017). For example, Reychav and Wu (2015) designed a mobile collaborative system in tablets by integrating group learning components for secondary students in geography classes. The collaborative learning activity in their study still preserves the advantages of face-to-face interactions, which are used to support rich communication (Reychav and Wu 2015). When engaging in collaborative learning, the task complexity, the information distribution among students, and group processes may influence the collective working memory of collaborators (Kirschner et al. 2018). Specifically, when group members have difficulties in exchanging knowledge and information due to improper task design, the collective working memory of the group will not be easy to form (Van Merriënboer and Sweller 2005; Kirschner et al. 2018). By influencing the information distribution and efforts of group members on information exchange, different modalities of using VM may have an impact on students' learning involvement during inquiry learning. For instance, whether team members could share the same working interface determines how much effort they need to put into information exchange. Besides, multiple working systems may also benefit groups, as group members could find multiple problem solutions simultaneously (Kirschner et al. 2018).

Cognitive load theory (Sweller et al. 1998; Van Merriënboer and Sweller 2005) presents a general framework intended to explain and predict how instructional methods and materials demand various learners' working memories. The assessment factors of cognitive load are mental load and mental effort (Paas and Van Merriënboer 1994). To be specific, mental load is associated with the number and degree of information interactions between task and subject. Mental effort is related to formats and manners of information presenting and teaching strategies (Paas and Van Merriënboer 1994). The mental effort level refers to the cognitive capacity required to complete a learning task, which could depict the mental involvement of students' during learning. Moreover, studies have shown that there is a complex relationship between learning performance and cognitive load, that is, the impact of cognitive load on learning performance are largely influenced by students' interaction with different learning materials (Wang et al. 2018). Cognitive systems could benefit from loading off working memory by making use of external resources (Pouw et al. 2014). Meanwhile, under collaborative learning, various interacting elements can be distributed among the working memories of other group members, thus reducing the cognitive load on single working memory (Kirschner et al. 2018).

Furthermore, Paas et al. (2005) proposed a Mental Effort-Performance coordinate system and a formula to compute task involvement of students when engaging in cognitive tasks. Calculating and representing the relative involvement of students in instructional conditions can provide a valuable

additive to research on the effectiveness of complex cognitive tasks (Paas et al. 2005). Besides, flow experience characterizes the subjective-environment interaction experience as playful and exploratory (Csikszentmihalyi 1990; Webster et al. 1993). Flow has been conceptualized as an optimal experience characterized by the perceived balance between challenge and skills (Ghani and Deshpande 1994). Studies have tested the positive effect on students' flow experience when using VM in science courses (e.g., Ibáñez et al. 2014). By investigating students' mental involvement under different learning situations, how students' learning engagement and experience vary could be discussed. Therefore, in this study, cognitive load, task involvement, and flow experience were involved in evaluating students' mental involvement during collaborative inquiry learning.

Purposes of This Study

The purpose of this study is to compare learning performance and mental involvement in the lens of different modalities of using VM for collaborative inquiry learning. Based on relevant empirical studies, we proposed three modalities of using VM for collaborative inquiry learning, specifically, one VM per student, one VM per group, and one VM per class. The study stresses not only learning performance but also the mental involvement of students when engaging in interactive collaborative learning environments. The learning performance of students was represented by scores of group work and subject test. The mental involvement of students was evaluated by cognitive load, task involvement, and flow experience. A designed collaborative learning activity was conducted in three primary science classes with these three modalities respectively. Two main research questions are described as follows:

1. What are the differences in learning performance among the three modalities of using VM in collaborative inquiry learning?
2. What are the differences in cognitive load, task involvement, and flow experience among the three modalities of using VM in collaborative inquiry learning?

Experiment Design

Virtual Lever Manipulative

The learning materials were developed according to the learning unit *Simple machines*, which is an important unit in science curriculums (3 to 6 grades). Lever, one of the classical simple machines in our daily life, is usually included in the *Simple machines* unit. In this study, we selected lever as the collaborative inquiry learning content and the inquiry topic is

about the law of lever. The learning objectives designed by researchers and the teacher in this study are as follows: identify fulcrum, load, and effort of a lever; explain the law of the lever by applying forces at different distances from the fulcrum. According to the science curriculum standards of Beijing, we developed a virtual lever manipulative (VLM) in the form of HTML 5, which can be accessed easily via browsers on tablets or computers. The fulcrum of the lever in VLM is in between the effort and the load. In VLM, two weights (each weighs 50 g) act as the load and a spring dynamometer works as the effort on the other side of the lever. Students can read the indications of the spring dynamometer when the lever balances. In VLM, there are five different positions of the load, 2.5, 5, 10, 15, and 20 cm, respectively. The position of the effort keeps 10 cm to the center of the lever. As shown in Fig. 1, there are two steps of the inquiry procedure. The first step is to rotate the nut to balance the lever. The second step is to change the position of the load and record the numerical value of effort shown in the spring dynamometer when the lever balances.

Participants

There were 80 fifth grade students who participated in this study and are all at the age of 10–12. They came from three science classes of the selected elementary school located in Beijing, China. All participants from the three classes are taught by the same science teacher. Most of the students had more than 5-year experience in using portable mobile devices; they all had a fair amount of collaborative learning experience before this experiment. Also, the results of the mid-term examination in this academic year showed that there was no significant difference among these three classes ($p = 0.273 > 0.05$), which means students of these three classes have a similar level of competence in science discipline. Students were randomly assigned to each group to work together since the beginning of this academic year. Each class was divided into eight groups and each group contains 5 to 6 students. Besides, all participants had prior experience using the tablet PCs provided in the science classroom of the school.

There are 28 students (12 boys, 16 girls) in class A, 26 students (14 boys, 12 girls) in class B, and 26 students (10 boys, 16 girls) in class C. The three classes were randomly assigned to three designed modalities of using VLM. To be specific, each of the students in class A had a VLM to interact with (one VLM per student); each group in class B had only one VLM to engage with (one VLM per group); students in class C recorded what the teacher showed and noted when their teacher is operating a VLM in front of the whole class (one VLM per class). Collaborative inquiry learning situations of the three classes are shown in Fig. 2.

Procedure

The experiment procedure is shown in Fig. 3. A week before the science collaborative learning activity, all students took a pre-test at the beginning of the experiment and the teacher conducted a prior knowledge teaching about the lever, such as fulcrum, load, and effort of a lever. During the learning activity, to facilitate face-to-face collaboration, students in each group were seated in a circle around a table in the science classroom. Each group was provided a group worksheet which prompted the inquiry procedure. They were expected to complete a group worksheet collaboratively. Specifically, each student of a group in class A operated his/her own VLM and conducted collaborative inquiry simultaneously; groups in class B had only one VLM to operate and discuss while groups in class C took notes on their notebooks according to what the teacher showed and taught when operating a VLM; after which, they exchanged information and finished the inquiry task. The VLM was run on provided tablet PCs for class A and class B. For class C, the teacher used a large projection screen in front of the classroom to show the operations of VLM. All the students in class C could observe the interactive process simultaneously. After completing the group worksheet, students were asked to complete a questionnaire survey on their cognitive load and flow experience. Finally, 1 week later, all participants took a post-test to evaluate their knowledge retention.

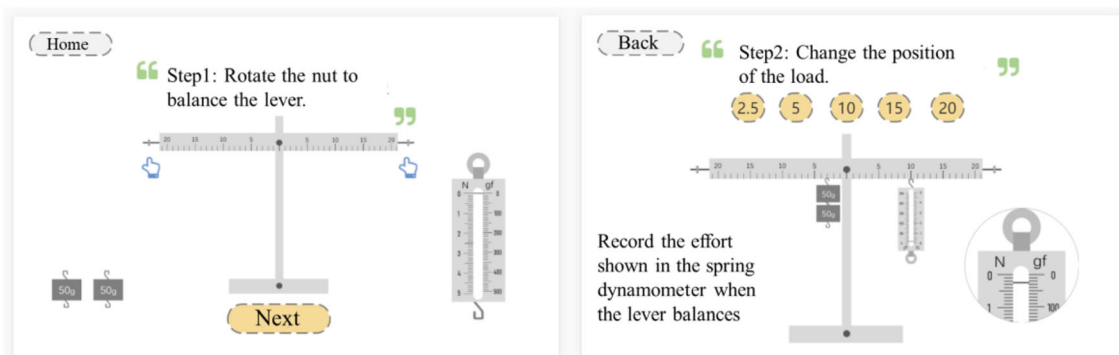


Fig. 1 Interfaces of virtual lever manipulative



Fig. 2 Collaborative situations of the three classes

Instruments

Assessment Tools

The pre-test aims to evaluate whether participants from these three classes had an equivalent prior knowledge before participating in this learning activity. It consists of six true or false questions and seven one-choice questions about the basic knowledge of lever with total scores of 30 and 70, respectively. During the learning task, students were asked to cooperatively complete a group worksheet. The group worksheet was designed to assess students’ learning outcome of the inquiry activity. There are two tasks in this group worksheet with total scores of 50 respectively. The first task is in the form of fill-in-a-table questions, which allow students to record the effort shown in the spring dynamometer when the lever balances under five different conditions. The second task is a discussion question to evaluate whether the students could draw the law of the lever correctly. Group members will receive the same grade according to their group worksheet. The post-test aims to evaluate students’ retention of relevant knowledge. It includes four fill-in-the-blank questions, four true or false questions, and two one-choice questions with total scores of 40, 40, and 20.

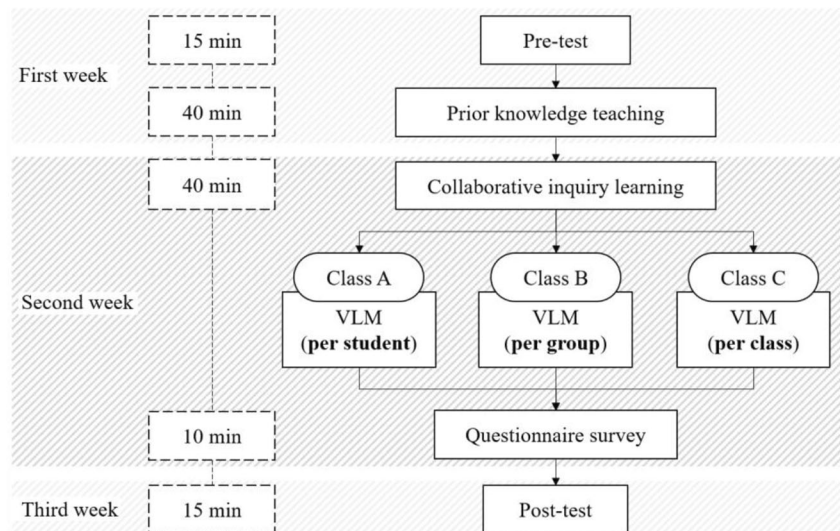
All questions in tests and the group worksheet were developed by three people: the classroom teacher according to the province’s science question bank, a teacher who had taught

the science course for more than 10 years, and an expert in science education were consulted to check these tests. Specific questions of the tests are different, but the contents covered by the tests are the same. The Cronbach’s alpha values of the pre-test and post-test are 0.715 and 0.706, respectively. As the group worksheet consists of only two inquiry tasks, the reliability value was not calculated.

Questionnaire

The questionnaire used after learning activities contains a cognitive load scale and a flow experience scale. Items of these scales are all in a six-point Likert rating. The cognitive load scale consists of mental load dimension and mental effort dimension with 5 items in each respectively. The scale was adapted from the research of Hwang et al. (2013) based on the measures proposed by Paas and Van Merriënboer (1994) and Sweller et al. (1998). The flow experience scale used in this study incorporates four dimensions: control, attention focus, curiosity, and intrinsic interests, with 3 items in each. This flow experience scale was developed by Park et al. (2010) based on the flow categories of Webster et al. (1993). The Cronbach’s alpha values of mental load (ML), mental effort (ME), control (CL), attention focus (AF), curiosity (CY), and intrinsic interests (II) are 0.884, 0.837, 0.625, 0.672, 0.916,

Fig. 3 Diagram of experiment design



and 0.710, respectively. This result means that all scales used in the study have acceptable reliability.

Results

Learning Performance

One-way ANOVA test for the three classes’ pre-test scores was not significant ($F(2,77) = 0.86, p = 0.426 > 0.05$). This means students in classes A, B, and C performed at a similar academic level in the pre-test, which could suggest that students in these three classes have a similar amount of prior knowledge before they engaged in learning materials and completing group worksheets. The mean and standard deviation values of group worksheet grades are shown in Table 1.

One-way ANOVA test showed that there was a significant difference between three classes on grades of group worksheets ($F(2, 77) = 3.52, p = 0.034 < 0.05$). Post hoc comparison turned out that class B performed significantly better than class A (mean difference = 15.30, $p = 0.031 < 0.05$), while the difference between class B and class C was not significant (mean difference = 4.81, $p = 0.71 > 0.05$), neither was the difference between class C and class A (mean difference = 10.50, $p = 0.186 > 0.05$). Paired-samples *t* test revealed that there were significant differences between pre-test and group worksheet for all the three classes (class A, $p < 0.001$; class B, $p < 0.001$; class C, $p < 0.001$). From the line chart (see Fig. 4), we could see a dramatic increase in all the three classes for the scores from pre-test to group worksheet. Class B (71.73) had a higher performance in group worksheet, followed by class C (66.92) and class A (56.43).

A week after the collaborative inquiry learning, students were arranged to take a post-test. The results of post-test are also shown in Table 1. One-way ANOVA test revealed statistically significant differences across three classes on post-test scores ($F(2, 77) = 8.22, p = 0.001 < 0.01$). Post hoc comparison

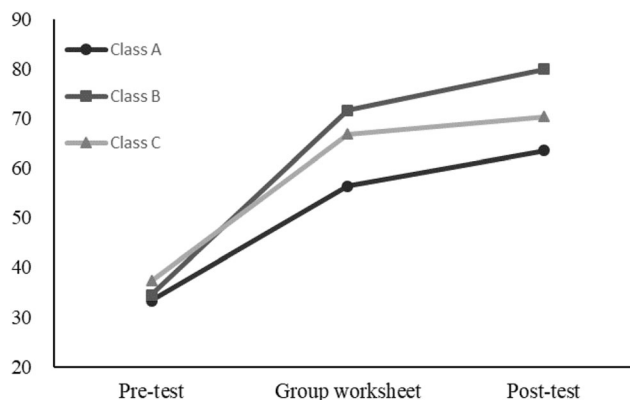


Fig. 4 Learning performance of the three classes

showed that class A continuously had a lower performance than class B (mean difference = - 16.43, $p < 0.001$), even though class A improved its score to 63.57. By contrast, in terms of knowledge retention in post-test, the difference between class B and class C was not significant (mean difference = 9.61, $p = 0.058 > 0.05$), neither was the difference between class C and class A (mean difference = 6.82, $p = 0.221 > 0.05$). Moreover, paired-samples *t* test showed that there was no significant difference between group worksheet and post-test for all the three classes (class A, $p = 0.068 > 0.05$; class B, $p = 0.165 > 0.05$; class C, $p = 0.396 > 0.05$).

It is worth noting that even though the difference between class B and class C on the scores of post-test was not statistically significant (mean difference = 9.61, $p = 0.058 > 0.05$), 0.058 is close to 0.05, the statistically significant level. Effect size can reveal the magnitude of the difference between groups (Cohen 1988). In this case, we calculated the effect size as a supplement for these classes from group worksheet to post-test. Cohen’s *d* formula (Cohen 1988) was adopted to determine the effect size for the experimental research by using means and standard deviations of groups. Cohen (1988) holds that when Cohen’s *d* is less than 0.2, the influence is relatively small; when Cohen’s *d* is about 0.5, the influence is moderate; and Cohen’s *d* higher than 0.8 means the difference is large. It is shown that Cohen’s *d* of the difference between class B and class A increased from 0.644 in group worksheet to 1.208 in post-test. For the difference between class B and class C, Cohen’s *d* also rose from 0.208 in group worksheet to 0.623 in post-test. Cohen’s *d* of the difference between class C and class A was 0.578 in group worksheet and 0.434 in post-test, respectively. These findings reveal that, in terms of learning performance, the gap between class B and class A increased 1 week later after the learning activity, so did the gap between class B and class C.

Cognitive Load

Cognitive load including mental load and mental effort was invested by the questionnaire immediately after the learning

Table 1 Descriptive analysis of learning performance in three classes

	Class	N	M	SD	F	p
Pre-test	A	28	33.27	11.04	0.86	0.426
	B	26	34.46	12.27		
	C	26	37.31	10.79		
Group worksheet	A	28	56.43	19.00	3.52*	0.034
	B	26	71.73	27.71		
	C	26	66.92	17.30		
Post-test	A	28	63.57	13.94	8.22**	0.001
	B	26	80.00	13.27		
	C	26	70.39	17.32		

* $p < 0.05$

** $p < 0.01$

activity. Mean and standard deviation values of mental load and mental effort are shown in Table 2.

One-way ANOVA test showed that there was no significant difference among three classes on mental load ($F(2, 77) = 0.95, p = 0.391 > 0.05$) and mental effort ($F(2, 77) = 0.20, p = 0.816 > 0.05$). Thus, there was no need to conduct a post hoc comparison. The result shows that even though there was no object to manipulate during the discussion session, students in class C did not report a significantly higher score on mental load and mental effort in comparison with the other two classes. This indicates that to analyze the experimental phenomena on VLM, students in all the three classes invested a similar amount of cognitive load into the learning tasks.

Task Involvement

Paas and his colleagues (2005) proposed the following formula to calculate task involvement (I) of each student involved in this study. R represents the z -score for mental effort, and P represents the z -score for performance.

$$I = \frac{R + P}{\sqrt{2}}$$

Post-test scores were used to represent students’ performance when calculating task involvement, as post-test scores refer to individual learning performance of the inquiry activities. By subtracting the total mean from each score and dividing the result by the total standard deviation, the students’ scores for mental effort and performance were standardized. The mean and standard deviation values of students’ task involvement are shown in Table 3.

One-way ANOVA test showed that there was a significant difference among the three classes on task involvement ($F(2,77) = 5.23, p = 0.007 < 0.01$). Post hoc comparison showed that the level of task involvement for class B was significantly higher than class A (mean difference = 0.82, $p = 0.005 < 0.01$). However, the difference between class B and class C (mean difference = 0.40, $p = 0.278 > 0.05$) and the difference between class C and class A (mean difference = 0.42, $p = 0.226 > 0.05$) on task involvement were not significant.

Table 2 Descriptive analysis of the cognitive load in three classes

	Class	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Mental load	A	28	2.18	0.93	0.95	0.391
	B	26	2.04	0.83		
	C	26	2.38	0.91		
Mental effort	A	28	2.97	0.99	0.20	0.816
	B	26	3.12	1.53		
	C	26	3.19	1.36		

Table 3 Descriptive analysis of the task involvement in three classes

	Class	<i>N</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Task involvement	A	28	-0.32	0.88	5.23**	0.007
	B	26	0.50	0.93		
	C	26	0.10	0.97		

** $p < 0.01$

Additionally, Mental Effort-Performance coordinate system can be used to visualize the motivational effects of instructional conditions (Paas et al. 2005). As shown in Fig. 5, each point in this coordinate system represents the mental effort z -score and related performance z -score of a participant in this study. A point above diagonal means the corresponding student had a relatively high involvement ($I > 0$), and vice versa. The percentages of students who had a relatively high involvement ($I > 0$) are 32.14% ($n = 9$) in class A. 65.38% of the students ($n = 17$) in class B while 53.85% of the students in class B ($n = 14$) showed a higher level ($I > 0$) of task involvement. The chi-square test showed that the students among the three classes had different task involvement levels, $\chi^2 = 6.187, p = 0.045 < 0.05$. These findings indicate that students in class B experienced the highest involvement during the collaborative task, while students from class A had a relatively low involvement. However, students of class C who just observed how the teacher manipulates VLM had a moderate level of task involvement.

Flow Experience

Flow experience including control, attention focus, curiosity, and intrinsic interests was also collected to explore students’

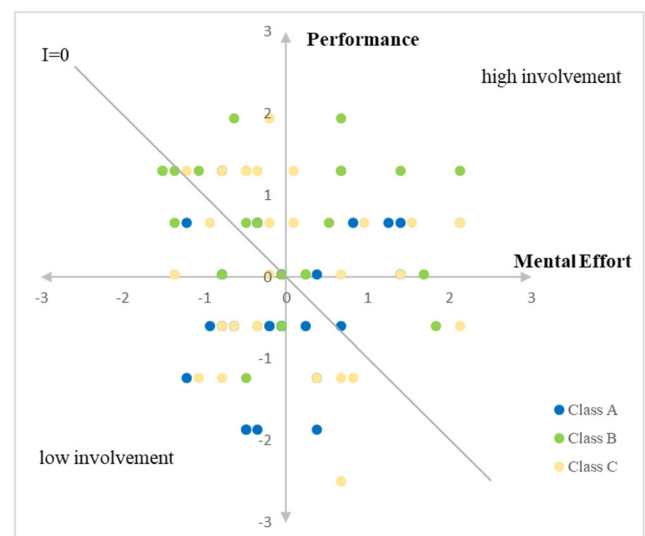


Fig. 5 Task involvement of students in three classes

learning engagement toward corresponding modalities of using VLM. The mean and standard deviation values of students' flow experience are shown in Table 4. One-way ANOVA test showed that the difference on "control" ($F(2,77) = 7.60, p = 0.001 < 0.01$), "attention focus" ($F(2,77) = 3.79, p = 0.027 < 0.05$), and "intrinsic interests" ($F(2,77) = 9.87, p < 0.001$) among three classes was significant. However, there was no significant difference on "curiosity" ($F(2,77) = 2.05, p = 0.136 > 0.05$), so we used post hoc comparison to get more details of the other three dimensions.

For "control" dimension, it is found that the level of "control" for class C was significantly lower than class A (mean difference = $-1.04, p = 0.001 < 0.01$) and class B (mean difference = $0.66, p = 0.046 < 0.05$). Students in class A had a VLM too at their own paces, while students in class B shared the same screen during the group interaction with VLM. By contrast with class C, in which there was no lever to operate, students in class A and class B had a higher sense of control in the course of the inquiry learning task. Furthermore, it is found that the difference between class A and class B was not significant (mean difference = $0.38, p = 0.351 > 0.05$). These findings may indicate that students could experience a higher level of "control" when they have opportunities to manipulate learning objects and master the process by themselves.

It is found that the level of "attention focus" for class A was significantly higher than that of class C (mean difference = $0.64, p = 0.038 < 0.05$). However, the difference between class A and class B (mean difference = $0.05, p = 0.976 > 0.05$) was not significant, neither was the difference between class B and class C (mean difference = $0.59, p = 0.068 > 0.05$). These findings may also reveal that the opportunities for manipulating learning objects could enhance the attention focus of students during inquiry learning.

Table 4 Descriptive analysis of the flow experience in three classes

	Class	N	M	SD	F	p
Control	A	28	4.80	0.94	7.60**	0.001
	B	26	4.42	1.01		
	C	26	3.76	1.02		
Attention focus	A	28	5.13	0.87	3.79*	0.027
	B	26	5.08	1.02		
	C	26	4.49	0.94		
Curiosity	A	28	4.68	1.47	2.05	0.136
	B	26	4.81	1.40		
	C	26	4.00	1.77		
Intrinsic interests	A	28	5.48	0.83	9.87***	0.000
	B	26	5.15	1.18		
	C	26	4.13	1.41		

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

As for "intrinsic interests," it turns out that class C was significantly lower than class A (mean difference = $-1.35, p < 0.001$) and class B (mean difference = $-1.02, p = 0.019 < 0.05$). Similarly, the difference between class A and class B on "intrinsic interests" was not significant (mean difference = $0.33, p = 0.580 > 0.05$). These results imply that students may experience a higher level of "intrinsic interests" if they could have more interactions with VLM, other than just observing.

Discussion and Conclusions

In this study, we proposed three modalities of using VLM for collaborative learning in a primary science course: one VLM per student, one VLM per group, and one VLM per class. We compared students' learning performance and mental involvement in a quasi-experimental manner in three classes. Through the collected data, this study depicts the difference in students' learning performance and mental involvement among these three modalities.

Under the assigned modalities of using VLM, groups in three classes experienced different affordances and formed different collaborative coordination. As mentioned by Larson and Christensen (1993), the formats of learning materials could influence the coordinate behaviors in a group. In class A (one VLM per student), students in groups sit around a table and could master the whole VM operating process by themselves. However, in order to work collaboratively, they had to check the operating stage and consistency of each other. In class B (one VLM per group), all students in groups were standing and clustering around the central VLM with a shared screen, even though they had fewer opportunities to operate the VLM. In class C (one VLM per class), groups experienced a traditional teaching method, with less movement during the discussion session. Students in class C only had their notebooks to refer to and could not do an extra test with a lever.

By analyzing the collected learning performance data, we could conclude that all the three classes had a dramatic increase in the scores from pre-test to group worksheet. It is also found that class B (one VLM per group) had significantly higher scores than class A (one VLM per student), indicating that in terms of the ways of using VLM, one VLM per student was less effective than one VLM per group. This might be because groups in class A wasted more time than class B on checking consistency with each other so that they had less time spending on the inquiry topic. The finding supports the view of Larson and Christensen (1993) by revealing that groups can benefit from improving their collective memory by increasing information exchange. Besides, there is no significant difference between class B (one VLM per group) and class C (one VLM per class) on post-test. The cause for this could be the familiarity of students with a given situation (Paas and Van Merriënboer 1994; Kirschner et al. 2018).

The modality of class C is more in line with traditional teaching methods, that is, observing the operation of the teacher and taking notes on what the teacher tells them about the experimental phenomenon. Moreover, students' perception of affordance is related to their relevant experience (Jeong and Hmelo-Silver 2016). Thus, students in class C might not have to spend much time regarding the instructional method that they were familiar with. Interestingly, a week later, the difference between class B and class C increased from a non-significant extent to an obvious extent. This indicates situations of class B and class C had nearly the same influence on immediate testing. However, concerning final knowledge retention, we could find that one VLM per group had a better effect than one VLM per class.

Meanwhile, there was no significant difference among the three classes on mental load and mental effort. This means that the formats and manners of information presenting the three modalities had no significant influence on students' cognitive load (Van Merriënboer and Sweller 2005), as the learning materials provided for all the classes had the same internal interactive process. However, the task involvement of students in class B was significantly higher than class A. In contrast, students in class C had a moderate level of task involvement. This finding implies the modalities of using VLM have an impact on task involvement of students during inquiry learning. Meanwhile, the applicability of the formula on task involvement proposed by Paas et al. (2005) in collaborative inquiry situations was verified in this study. As for the flow experience, it was found that the sense of control and intrinsic interests of class C were significantly lower than those of class A and class B. Moreover, class A's attention focus was also higher than class C, while the differences between class A and class B on the four dimensions of flow experience were not obvious. This might be because students in class C had the lowest autonomy and fewest interactions with learning materials so that they seldom had playful and exploratory feelings during the inquiry process (Csikszentmihalyi 1990; Webster et al. 1993). In this circumstance, it could be concluded that the opportunities for operating VLM could enhance the flow experience of students during inquiry learning. Hands-on experience of students needs to be considered when conducting instructional design for science inquiry learning. However, compared with flow experience, the task involvement of students might have a stronger influence on learning performance. In addition, the finding that class B (one VLM per group) had superior learning performance than class A (one VLM per student) is consistent with the finding of Lin et al. (2012). It should also be noted that, however, the results of this research are not completely consistent with the results of Lin et al. (2012). Specifically, the perceptions toward learning experience were not significantly different between one VLM per student and one VLM per group, even though the task involvement varies obviously.

The value of this study is to compare the learning performance and mental involvement of the proposed three modalities of using VM for groups which have not been fully discussed in former VM-related studies. The traditional teaching method in which the teacher shows the operation of VM on a projection screen was discussed in this study; student-mobile device ratios of collaborative inquiry learning under the support of mobile technologies were explored as well. Meanwhile, we concerned not only the influence on conceptual understanding but also the mental engagement and perception during the collaborative inquiry learning process. This could unfold a real landscape of the perception and experience of students with given constraints in learning tasks. Meanwhile, the findings of learning performances and mental involvement shown in this study could give a lens of instructional scaffolds design to afford active and effective face-to-face collaborative learning (Kirschner et al. 2018). In terms of limitations in this study, the multimodal analysis, which stresses on gestures and verbal communication, was not conducted due to the quality issues of recorded videos. Besides, Cronbach's alpha values for control and attention focus are relatively low (< 0.70); thus, we need to treat relevant results carefully. In future work, we tend to invest how group characteristics mediate interactions during collaborative learning with these three modalities of using VM. A multimodal analysis method would be used to analyze the interaction process. Meanwhile, we would concern the long-term collaborative coordination pattern of groups when provided specific ways of using VM in science courses.

Funding Information This work was supported by the China Scholarship Council (201806010278).

Compliance with Ethical Standards

Prior to formal experiments, all the participants were informed to sign the written consents for this study; All participants were informed the right of voluntary participation and the freedom of quitting the study.

Conflict of Interest The authors declare that they have no conflict of interest.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: models, tools, and challenges. *International Journal of Science*, 32(3), 349–377.
- Bouck, E. C., & Flanagan, S. M. (2009). Virtual manipulatives. *Intervention in School and Clinic*, 45(3), 186–191.
- Cáceres, M., Nussbaum, M., Marroquín, M., Gleisner, S., & Marquinez, J. T. (2018). Building arguments: key to collaborative scaffolding. *Interactive Learning Environments*, 26(3), 355–371.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). New Jersey: L. Erlbaum Associates.
- Crompton, H., Burke, D., Gregory, K. H., & Gräbe, C. (2016). The use of mobile learning in science: a systematic review. *Journal of Science Education and Technology*, 25(2), 149–160.
- Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers in Education*, 110, 51–63.
- Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. New York: HarperCollins Publishers.
- Ghani, J. A., & Deshpande, S. P. (1994). Task characteristics and the experience of optimal flow in human—computer interaction. *The Journal of Psychology*, 128(4), 381–391.
- Ha, O., & Fang, N. (2017). Interactive virtual and physical manipulatives for improving students' spatial skills. *Journal of Educational Computing Research*, 55(8), 1088–1110.
- Hwang, G. J., Yang, L. H., & Wang, S. Y. (2013). A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, 69, 121–130.
- 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 in Education*, 71(2), 1–13.
- 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.
- Kapici, H. O., Akcay, H., & de Jong, T. (2019). Using hands-on and virtual laboratories alone or together—which works better for acquiring knowledge and skills? *Journal of Science Education and Technology*, 28(3), 231–250.
- 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.
- Koning, B. B. D., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualizations: an embodied perspective. *Educational Psychology Review*, 23(4), 501–521.
- Larson, J. R., & Christensen, C. (1993). Groups as problem-solving units: toward a new meaning of social cognition. *The British Journal of Social Psychology*, 32(1), 5–30.
- 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.
- 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.
- Min, J., Lin, Y. T., & Tsai, H. C. (2016). Mobile APP for motivation to learning: an engineering case. *Interactive Learning Environments*, 24(8), 2048–2057.
- Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics*, 47(2), 175–197.
- Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics*, 8(6), 372–377.
- Mulet, J., van de Leemput, C., & Amadieu, F. (2019). A critical literature review of perceptions of tablets for learning in primary and secondary schools. *Educational Psychology Review*, 31, 1–32.
- 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., & Van Merriënboer, J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351–371.
- Paas, F., Tuovinen, J. E., Jeroen, J. G. V. M., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: optimizing learner involvement in instruction. *Educational Technology Research and Development*, 53(3), 25–34.
- Park, J., Parsons, D., & Ryu, H. (2010). To flow and not to freeze: applying flow experience to mobile learning. *IEEE Transactions on Learning Technologies*, 3(1), 56–67.
- Pouw, W. T. J. L., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, 26(1), 51–72.
- Reychav, I., & Wu, D. (2015). Mobile collaborative learning: the role of individual learning in groups through text and video content delivery in tablets. *Computers in Human Behavior*, 50, 520–534.
- Sweller, J., Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review*, 17(2), 147–177.
- Wang, T., & Tseng, Y. (2018). The comparative effectiveness of physical, virtual, and virtual-physical manipulatives on third-grade students' science achievement and conceptual understanding of evaporation and condensation. *International Journal of Science and Mathematics Education*, 16(2), 203–219.
- Wang, C., Fang, T., & Miao, R. (2018). Learning performance and cognitive load in mobile learning: impact of interaction complexity. *Journal of Computer Assisted Learning*, 34(6), 917–927.
- Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, 9(4), 411–426.
- Wong, L., & Looi, C. (2011). What seems do we remove in mobile-assisted seamless learning? A critical review of the literature. *Computers in Education*, 57(4), 2364–2381.
- Yuan, Y., Lee, C. Y., & Wang, C. H. (2010). A comparison study of polyominoes explorations in a physical and virtual manipulative environment. *Journal of Computer Assisted Learning*, 26(4), 307–316.
- Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, 21(3), 317–331.
- Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching*, 45(9), 1021–1035.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.