Science Education

Blending Physical and Virtual Manipulatives: An Effort to Improve Students' Conceptual Understanding Through Science Laboratory Experimentation

GEORGIOS OLYMPIOU, ZACHARIAS C. ZACHARIA
Learning in Science Group, Department of Educational Sciences, University of Cyprus, P. O. Box 20537, Nicosia 1678, Cyprus

ABSTRACT: This study aimed to investigate the effect of experimenting with physical manipulatives (PM), virtual manipulatives (VM), and a blended combination of PM and VM on undergraduate students' understanding of concepts in the domain of *Light and Color*. A pre–post comparison study design was used for the purposes of this study that involved 70 participants assigned to three conditions. The first condition consisted of 23 students that used PM, the second condition consisted of 23 students that used VM, and the third condition consisted of 24 students that used the blended combination of PM and VM. In the case of the blended combination, the use of VM or PM was selected based on whether it provides an affordance that the other medium of experimentation (PM or VM) cannot provide. All conditions used the same inquiry-oriented curriculum materials and procedures. Conceptual tests were administered to assess students' understanding before, during, and after teaching. Results revealed that the use of a blended combination of PM and VM enhanced students' conceptual understanding in the domain of *Light and Color* more than the use of PM or VM alone. © 2011 Wiley Periodicals, Inc. *Sci Ed* **96:**21–47, 2012

INTRODUCTION

Over the past few decades, several research studies have attempted to investigate and document the value of using physical manipulatives (PM) (real-world physical/concrete material and apparatus) and virtual manipulatives (VM) (virtual apparatus and material, which exist in virtual environments, such as computer-based simulations) in science laboratory experimentation (Balamuralithara & Woods, 2009; Hofstein & Lunetta, 2004; Jaakkola, Nurmi, & Veermans, 2010; Toth, Morrow, & Ludvico, 2009; Triona & Klahr, 2003; Winn et al., 2006; Zacharia, 2005; Zacharia & Anderson, 2003; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011; Zacharia, Olympiou, & Papaevripidou, 2008). Comparative studies have been undertaken to identify which of these two modes of experimentation (PM or VM) is the most preferable across several science subject domains (Finkelstein et al., 2005; Klahr, Triona, & Williams, 2007; Mosterman et al., 1994; Toth et al., 2009; Triona & Klahr, 2003; Zacharia, 2007; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011; Zacharia et al., 2008). Findings from these studies revealed instances where the use of VM would appear to be as beneficial to student learning as PM (Klahr et al., 2007; Triona & Klahr, 2003, Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011), more beneficial to student learning than the use of PM (Finkelstein et al., 2005; Zacharia, 2007; Zacharia et al., 2008) and vice versa (Gire et al., 2010; Marshall & Young, 2006).

A question that is raised at this point is why the findings of these studies appear to be inconsistent with each other. A comparison across the material and methods used in these studies revealed that the differences in outcomes were caused primarily by the differing affordances that the PM and VM from each study carried (Zacharia et al., 2008). By affordances, we mean the qualities of PM or VM that offer the possibility of an interaction relative to the ability of a learner to interact. Both PM and VM carry numerous affordances that overlap significantly (e.g., both of them could provide students with the opportunity to set up and run a lab experiment). On the other hand, PM and VM carry a number of different affordances, which were found to provide students with unique learning experiences (Finkelstein et al., 2005; Winn et al., 2006; Zacharia 2007; Zacharia et al., 2008). For instance, only PM can offer students experiences that involve the manipulation of the actual items of a lab experiment (perceptual-motor skills). Conversely, only VM can provide students with opportunities to manipulate the conceptual objects involved in a lab experiment (objects that have no perceptual fidelity).

Given these differing affordances, a number of researchers have advocated in favor of combining the use of PM and VM (Campbell, Bourne, Mosterman, & Brodersen, 2002; Jaakkola & Nurmi, 2008; Jaakkola et al., 2010; Toth et al., 2009; Winn et al., 2006; Yueh & Sheen, 2009; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011; Zacharia et al., 2008), because it is the only way to reap and use the benefits (advantageous affordances) that both PM and VM carry (Winn et al., 2006; Zacharia et al., 2008). However, there is no framework available at the moment that portrays how these affordances can be used to combine PM and VM for the purposes of science experimentation (Zacharia et al., 2008).

Science education research so far, has focused primarily on PM and VM sequential combinations, in which PM and VM were used in an alternating manner for the same (e.g., Jaakkola & Nurmi, 2008; Jaakkola et al., 2010) or different content/experiments (e.g., Gire et al., 2010; Toth et al., 2009; Winn et al., 2006; Zacharia 2007; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011; Zacharia et al., 2008). These sequential combinations were formed based upon the methodological needs of the study (e.g., providing equal opportunities for both PM and VM or whether students using PM can switch to using VM

and vice versa), rather than on matching the intervention to the needs of the content of a study (e.g., to the goals of each experiment). As a result, there is no information coming from this research domain on whether a more well thought out PM and VM combination, which takes into consideration the PM and VM affordances and specifically targets the content of each lab experiment separately, could enhance students' learning more than the use of PM or VM alone.

The purpose of this study was to contribute to this direction, namely to the development and implementation of a framework that portrays how PM and VM could be blended on the basis of their affordances to better serve the learning goals of a lab experiment compared to using PM or VM alone. In this study, we use the terms "blend" and "blended combination" in an attempt to separate this research effort from prior research efforts that involved sequential combinations. By blended combinations, we imply that the PM and VM mixing is based upon a framework that intentionally addresses the needs of the content/goals of each lab experiment separately, rather than assigning PM or VM to different sets of lab experiments as in sequential combinations (e.g., students conduct the first set of the experiments of a certain curriculum with PM and the other set with VM). In the case of blended combinations, whenever possible, the PM and VM are combined and used in conjunction in the context of each experiment in a way that they match the needs of each experiment separately. Whereas in the case of the sequential combinations, only PM or VM are used for different sets of experiments without receiving any support from one another at any time. It should be noted at this point that the term blended combination was also introduced in prior science education studies that focused on combining PM and VM (e.g., Toth et al., 2009), but it was defined in the manner in which we define sequential combinations. Hence, the findings of these studies are not of a same nature as the findings of this study.

Returning to the purpose of this study, we specifically set as our overarching learning goal the improvement of students' understanding of concepts in the physics domain on *Light & Color* and proceeded with the development of a framework that blends PM and VM according to the impact their unique affordances have on students' conceptual understanding. In doing so, we identified through prior research, the PM and VM unique affordances that were found to support students' conceptual understanding across several science domains (e.g., Hsu, 2008; Zacharia et al., 2008) and created blended combinations of PM and VM for each of the study's experiments. To test the effectiveness of these combinations, we designed and ran this research (experimental) study, in which the use of the PM and VM blended combinations was compared to the uses of only PM or VM. The research question of this study was, Should the use of a blended combination of PM and VM, as described above, be preferred over the use of PM or VM alone when the enhancement of students' conceptual understanding through laboratory experimentation in the domain of Light & Color is at task? Given the aforementioned targeted selection of PM and VM for creating the blended combinations used in this study, it was hypothesized that the use of the blended combinations would enhance students' understanding of scientific concepts more than the use of only PM or VM.

THEORETICAL BACKGROUND

PM and VM Affordances

The literature on PM and VM laboratory experimentation in science education has highlighted the fact that PM and VM have a significant overlap in terms of the affordances

that they could offer in laboratory experimentation (e.g., manipulation of material) and that under certain conditions they could have a similar effect on students' conceptual understanding (Triona & Klahr, 2003; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011). For instance, the use of both VM and PM could provide a perceptual grounding for concepts that might otherwise be too abstract to be easily understood (Winn et al., 2006); promote an active, "hands-on," problem-solving stance that, in turn, often fosters a deep understanding of a phenomenon (Triona & Klahr, 2003); and provide effective exposure to experimentation and its corresponding skills (Hofstein & Lunetta, 2004). On the other hand, the literature of this domain revealed that PM and VM carry certain (unique) affordances that differ between them. Thus, their presence during laboratory experimentation results in a different effect on student learning, in favor of the manipulatives that carry these additional affordances/advantages (Gire et al., 2010; Finkelstein et al., 2005; Winn et al., 2006; Zacharia, 2007; Zacharia et al., 2008).

In the case of PM, physicality (actual and active touch of concrete material) is reported as one such unique affordance (see, e.g., Feisel & Rosa, 2005). Research studies focusing on physicality and its impact on learning provide evidence that physicality forms the basis for conscious memory and learning (Bara, Gentaz, Pascale, & Sprenger-Charolles, 2004; Klatzky & Lederman, 2002; Loomis & Lederman, 1986). According to Gire et al. (2010), physical, hands-on science investigations allow students to experience science phenomena directly through experimentation with physical materials and by designing and engineering physical artifacts. They further argue that through these processes, students can gain experience in planning investigations, using appropriate scientific instruments, and collecting, recording, and analyzing real-world data.

A second beneficial affordance of PM is that measurement errors are naturally present, whereas in VM measurement errors are often ignored. Competency in a domain includes the knowledge that various types of measurement errors exist and the ability to deal with them (Toth, Klahr, & Chen, 2000). Reading instruments in VM, for example (even with a possibility to zoom in), is by its nature easier than using and reading physical/concrete instruments. Maisch, Ney, van Joolingen, and de Jong (2009) showed that knowledge concerning measurement errors, acquired in a nonexperimental context, does not transfer easily to the students' actions in a PM laboratory, which suggests that PM play a specific role. In this context, Chen (2010) recently asserted that VM often display a too idealized world, leading to a limited view on experimentation.

In addition to these affordances, Balamuralithara and Woods (2009) list 13 objectives for the use of PM, which can be summarized in terms of three affordances, namely acquisition of psychomotor skills, awareness of safety procedures, and learning how to use human senses for observations (for more details on PM also see Hofstein & Lunetta, 2004). Alternatively, the use of VM, unlike the use of PM, could (a) provide capabilities for altering the natural time scale and simplifying real-world models, thus making phenomena more visible to learners, thereby accommodating individual cognitive levels (deJong & Njoo, 1992), (b) provide an information-rich and multiple representation (verbal, numerical, pictorial, conceptual, and graphical) environment (Hsu & Thomas, 2002), (c) allow students to change variables which would be impossible or unrealistic to change in the natural world (e.g., global temperature, a person's blood pressure) (Windschitl, 2000), (d) provide immediate feedback about errors to the students and thus the opportunity to repeat the same experiment immediately (Huppert & Lazarowitz, 2002; Ronen & Eliahu, 2000), (e) facilitate learning by focusing students' attention more directly on the targeted phenomena (deJong & Van Joolingen, 1998), (f) allow students to visualize objects and processes that are normally beyond perception (Winn et al., 2006) and to simplify them

(Hsu, 2008), (g) allow students to perform a wide range of experiments faster and more easily and thus experience more examples (Carlsen & Andre, 1992; Huppert & Lazarowitz, 2002), (h) enable students to experience what might be too expensive or difficult to carry out with PM and permit experiments to be performed repeatedly in a safe environment (Doerr, 1997; Faryniarz & Lockwood, 1992), and (i) provide scaffolds, which are tools, either cognitive (Jonassen, 2000) or social (Tabak & Baumgartner, 2004), enabling students to perform the processes (e.g., inquiry processes; Kim, Hannafin, & Bryan, 2007) they would not be able to perform competently without the tool's support (Salomon, Perkins, & Globerson, 1991; for a review, see deJong, 2006).

Given all these affordances, it becomes obvious that VM carry many more unique affordances than PM. The reason behind this is that VM emerged to address the need to complement PM which presented a number of inherent deficiencies within the context of school science experimentation. The attempt was for VM to "match" the experimental affordances provided by PM and to exceed them by providing even more affordances than PM. As a result, there is a number of VM across science subject domains providing representations that appear to be just as personally meaningful to students as PM and even more manageable, "clean," flexible, and extensible than their physical counterparts (Triona & Klahr, 2003). Hence, the use of VM, unlike the use of PM, could provide affordances, such as portability, safety, cost-efficiency, scaffolding, minimization of errors, amplification or reduction of temporal and spatial dimensions, manipulation of reified objects, and flexible, rapid, and dynamic data displays (Hsu & Thomas, 2002).

A number of studies involving the use of PM and VM (which carry a number of the aforementioned advantageous affordances) showed both that the use of VM enhanced student learning more than the use of PM and vice versa (Gire et al., 2010; Finkelstein et al., 2005; Marshall & Young, 2006). For instance, Finkelstein et al. (2005) compared two groups of students, those who used PM and those who used VM (a computer simulation) that explicitly modeled electron flow. The comparison focused on students' understanding of physics concepts and skills with real equipment. The findings showed that students who used the simulated equipment carrying the additional affordance/advantage outperformed their counterparts both on a conceptual survey of the domain and in the coordinated tasks of assembling a real circuit and describing how it worked.

Conversely, through their study, Gire et al. (2010) revealed instances where PM were more advantageous for student learning than VM. More specifically, they found that whenever a grounded physical experience was involved, such as examining the lifting of an object with a movable pulley compared to a fixed pulley, PM appeared to enhance students' conceptual understanding more than VM. According to Gire et al., PM experimentation was found to have an advantage over the VM experimentation because PM carried the affordance of physicality (touch sensory input), which was apparently necessary to understand the pulley-related concepts introduced through the experiments. In fact, it was found that the haptic feedback acquired in using pulleys supported students' understanding of the concept of force in the pulley domain.

Finally, it is important to note that the aforementioned PM and VM unique affordances were found to be conducive to students' conceptual understanding across several subject domains and age groups. For example, the VM affordance of the manipulation of reified objects was found to have a positive effect on students' conceptual understanding in chemistry (e.g., Martínez-Jiménez, Pones-Pedrajas, Climent-Bellido, & Polo, 2003; Wu, Krajcik, & Soloway, 2001), physics (e.g., Finkelstein et al., 2005; Zacharia, 2007), and biology (e.g., Huppert & Lazarowitz, 2002; Toth et al., 2009). Someone could reasonably argue that these

affordances are content independent, which provides us with the opportunity to use this knowledge across the science subject domains. On the other hand, research has shown that these affordances are learning-objective dependent. In other words, research has associated all these affordances with specific learning objectives (Zacharia et al., 2008). This means that PM or VM should be used when their affordances, unique or not, serve the objectives of an experiment.

PM and VM Experimentation and Conceptual Understanding

Over the years, researchers have shown that conceptual understanding in science education is facilitated through learning that promotes conceptual change (Carey & Spelke, 1994; Chi, Slotta, & deLeeuw, 1994; diSessa, 2008; Limon & Mason, 2002; Piaget, 1985; Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou, Vamvakoussi, & Skopeliti, 2008). Piaget (1985) argued that, to foster conceptual change, students have to be confronted with "discrepant events" that contradict their conceptions and invoke a "disequilibration or cognitive conflict" that puts students in a state of reflection and resolution. Research has shown that these discrepant events could be provided effectively through the PM or VM inquiry-based science experimentation (Hofstein & Lunetta, 2004; Tao & Gunstone, 1999; Zacharia & Anderson, 2003; Zacharia et al., 2008).

PM and VM through their affordances offer students the possibility to inquire into the event presented, to alter the values of parameters, to initiate processes, to probe conditions, and to observe the results of these actions. In this way, students can interpret the underlying scientific conceptions of the PM or VM experiment, compare these with their own conceptions, formulate and test hypotheses, and reconcile any discrepancy (resulted through cognitive conflict) between their ideas and the observations from the experiment (Tao & Gunstone, 1999). Prior research suggests that cognitive conflict models of instruction, such as inquiry-based experimentation, can be effective in promoting conceptual change and thus, in enhancing students' conceptual understanding (for details, see Guzzetti, Snyder, Glass, & Gamas, 1993).

In addition to promoting conceptual understanding through conceptual change, PM and VM experimentation also facilitates conceptual understanding through framing the content in terms of its *experiential value* and by scaffolding *reseeing*. Framing content in terms of *experiential value* refers to the act of emphasizing the potential value that the content has to enrich and expand students' everyday experience. PM and VM inquiry-based experimentation have proven to provide such framing that positions students in a meaningful learning environment (Hofstein & Lunetta, 2004; National Research Council, 1996), which portrays the experiential value of content by illustrating its immediate usefulness in everyday life (Girod & Wong, 2002). Engle (2006) illustrated the important role that framing plays in shaping the way that students engage with content and how this type of engagement leads to learning.

Reseeing refers to the act of going beyond one's current perceptions of everyday objects and viewing them through the lens of a new idea (Girod, Rau, & Schepige, 2003). PM and VM experimentation can scaffold *reseeing* through their affordances. For instance, when using PM, a student can *resee* an object's characteristics (e.g., hardness, texture, weight, inertia, geometry/shape, smoothness, slippage, temperature) through tactual/haptic sensation (Loomis & Lederman, 1986), whereas in the case of VM, a student can *resee* the same object through multiple representations (e.g., through concrete representations that are then faded into more idealized ones; see for more details Goldstone & Son, 2005).

Such PM or VM *reseeing* contexts can enrich students' knowledge about an object or a phenomenon and thus further refine their conceptual understanding. In addition, such contexts can make objects or phenomena more visible to learners thereby accommodating individual cognitive levels (deJong & Njoo, 1992).

Combining PM and VM: Findings From Prior Research

PM and VM experimentation have traditionally been considered competing methods in science classrooms (Jaakkola & Nurmi, 2008). However, after recognizing the unique affordances that PM and VM carry, researchers have begun to explore (a) whether it is possible for PM and VM to be combined and (b) the potential benefits of combining PM and VM experimentation rather than using them alone. The studies that investigated whether PM and VM can coexist in a learning environment revealed that it is possible to combine the two. For example, Zacharia and colleagues, through a number of studies across several subject domains in physics (e.g., Zacharia & Anderson, 2003; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011), investigated whether the transition from one mode of experimentation to the other (PM to VM and vice versa) is feasible, given that the nature of motor skills involved in physical and virtual manipulation is different. Their findings showed that the transition is feasible and that it does not have a different effect on students' conceptual understanding.

These studies also showed that it is possible to start a sequence of experiments with either PM or VM. Winn et al. (2006) showed that PM should precede VM when there is a need to contextualize learning for students with little prior experience of the phenomenon or system under study (e.g., the study of ocean currents; for more details, see Winn et al., 2006). On the other hand, Zacharia and Anderson (2003) found that VM should precede PM when the PM experimentation concerns a rather complex phenomenon or system. In such a case, a VM of low fidelity is used, where these omit details found in PM and focus only on the to-be-learned structural features (Zacharia & Anderson, 2003).

In the case of studies focusing on whether PM and VM combinations should be preferred over the use of PM or VM alone, it was found that use of PM and VM combinations is more beneficial. For example, findings showed that the use of VM in conjunction with PM resulted in an improvement in learning of abstract physical phenomena, helping students construct mental models that explain observable results of PM experiments (Zollman, Rebello, & Hogg, 2002). Akpan and Andre (2000) investigated the learning of a skill (the dissection of a frog) and found that students who worked only with VM or with VM preceding actual dissection (PM) outperformed students performing only the hands-on dissection (PM) on a test measuring the knowledge of frog anatomy. Martínez-Jiménez et al. (2003) working within the domain of chemistry (e.g., distillation) compared students who learned through PM only, with students who used VM preceding the PM. Their results showed that students reached higher levels of conceptual understanding in the combination group. Zacharia (2007) had two groups of students explore electrical circuits. One group followed the curriculum using a PM, whereas the other group began with the use of VM and moved to the use of PM half way through the course. The students who used the PM and VM combination were found to have better conceptual understanding than the group that used only PM. The combination group also displayed an advantage half way through the course, which indicates that the virtual laboratory better promoted the acquisition of conceptual knowledge. The advantage gained through the use of VM was obviously taken over in the part in which both groups used PM. Similar findings to the Zacharia (2007) study were found in the Zacharia et al. (2008) study, but in a different subject domain (heat

and temperature). The fact that the findings of these two studies were the same implies that they are not subject domain dependent.

Jaakkola and Nurmi (2008) worked with students who had to complete assignments on electrical circuits. They had three conditions, a VM lab, a PM lab, and a condition in which students first had to work through the VM and then do the same assignments with PM. They found the combined condition was the most advantageous for acquiring conceptual knowledge, followed by the VM, with the PM condition yielding the lowest scores. Toth et al. (2009) compared sequences of VM before PM, with the reverse sequence (PM before VM) for the domain of DNA gel electrophoresis and reported a small advantage for the VM first group on a conceptual knowledge test, but the difference was not significant.

Overall, research so far has shown that combining PM and VM is more beneficial than using them alone. However, it has not provided us with a framework as to how to optimize the effect of these combinations. In fact, the only combinations used so far were sequential in nature and were not supported by any framework (e.g., Gire et al., 2010; Jaakkola & Nurmi, 2008; Jaakkola et al., 2010; Toth et al., 2009; Winn et al., 2006; Zacharia 2007; Zacharia et al., 2008; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011). The studies in this domain were rather focusing to serve the methodological needs that each study had. As a result, there is no information emerging from this research domain on whether a more well thought out PM and VM combination, taking into consideration the PM and VM affordances and specifically targeting the content of each lab experiment separately, could enhance students' learning more than the use of PM or VM alone.

A Framework for Blending PM and VM

Zacharia et al. (2008) suggested that the best way to develop a framework that portrays how PM and VM should be blended, to enhance students' conceptual understanding more than when PM and VM are used alone, is to take the learning objectives of an experiment/activity and carefully analyze them in terms of what the student should be introduced/exposed to (e.g., an authentic real experience; an experience of measurement errors; an experience that involves observation of reified objects, such as atoms). Given this analysis, PM and VM should be blended in a way that best serves what has been identified as important for the students to experience. In other words, using the learning objectives of each experiment as the criterion for blending PM and VM enables us to tailor the use of PM or VM to what is required by each experiment for the students to experience in the best possible way. This implies that the pedagogical and didactical parameters of an experiment (e.g., content, collaboration, cognitive skills), which are reflected through its learning objectives, are better served through a research-based, targeted use of PM and VM. The only "drawback" of this framework is that for its development presupposes knowledge of what PM and VM could offer, particularly, in terms of unique affordances.

For the purposes of this study, we have taken the suggested Zacharia et al (2008) ideas, transformed them into a framework (see Figure 1) and implemented it. Right below, we describe our reasoning and actions while following the steps outlined in our framework for the purposes of creating the PM and VM blends of this study. This aims at situating the framework in a real context.

First, we took the study's curriculum/teaching material and identified the overarching general learning objective (the promotion of conceptual understanding) and then
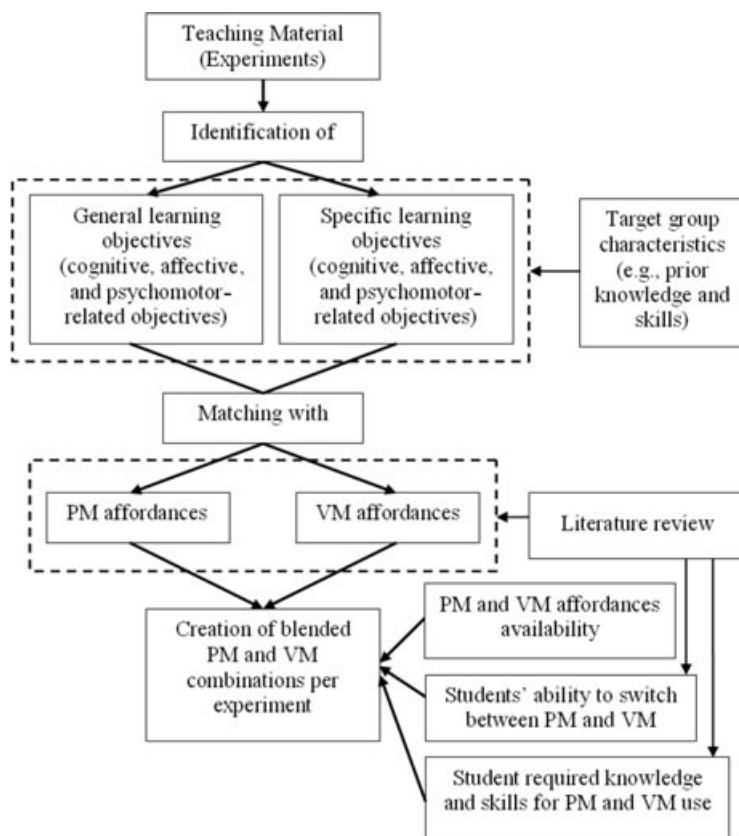
Figure 1. The study's framework for blending PM and VM.

for each one of the teaching material's experiments we identified their specific learning objectives. There was no need to make any adjustments to the learning objectives according to our participants' characteristics (e.g., prior knowledge and skills) because the teaching material used in our study was designed for college students. Moreover, the selected teaching material does not require any prior knowledge of the domain, as it starts with experiments that introduce the very first/basic concepts of the domain of Light & Color.

Second, we reviewed the relevant literature and identified PM and VM unique affordances that were found to promote the learning objectives identified in the first step. Third, we matched the learning objectives with the corresponding PM and VM unique affordances. In Table 1, we provide a sample of such identified learning objectives that were matched to PM or VM unique affordances. For example, whenever one of the experiments' objectives involved taking and analyzing numerous accurate measurements (for the purposes of reaching to valid conclusions, a certain type of graph or formula etc.), VM was preferred over PM because they were the only manipulatives that could provide a measurement error-free learning environment. PM use always involve measurement errors (Maisch et al., 2009; Toth et al., 2000), which were found to be distractive to student learning, especially when students are introduced to complex concepts/phenomena or concepts/phenomena for the first time (Zacharia & Constantinou, 2008; Zacharia et al.,2008). On the other hand,

TABLE 1
A Sample of the Study's Specific Learning Objectives and the PM and VM Unique Affordances That Were Matched With Them

Learning Objective	Type of Manipulative	Affordance	Reference
Observing the real phenomenon	PM	Provision of an authentic-concrete experience	Gire et al. (2010)
Experiencing certain characteristics of a concrete object	PM	Sensing an objects' roughness, viscosity, etc.	Loomis and Lederman (1986)
Experiencing certain motor skills needed for handling certain material and apparatus	PM	Use of concrete material and apparatus	Triona and Klahr, 2003
Observing phenomena that cannot be observed in real life	VM	Provision of representations of reified objects (e.g., light rays)	Finkelstein et al. (2005)
Making or repeating accurate measurements	VM	No measurement errors	Zacharia et al. (2008)
Making or repeating measurements quickly	VM	Overcoming time consuming procedures	Zacharia et al. (2008)

whenever the objective was to show to the students that measurement errors are naturally present in science experimentation and that competency in a domain includes the knowledge that various types of measurement errors exist and the ability to deal with them, PM was preferred over VM because in VM measurement errors are often ignored. Recent research showed that when VM is idealized (measurement error-free learning environment), students tend to get a limited view on experimentation (Chen, 2010). The fact that neither PM nor VM could provide both of the measurement error related affordances also portrays the essense of blending PM and VM. In other words, a student in a blended PM and VM condition could be introduced both to the "messy" nature of science and to studying a phenomenon without unnecessary distractions.

Fourth, we examined whether the required affordances were available through the PM and VM that we had access to. In the case of PM, all of the required affordances were available, whereas in the case of VM, they were not. Thus, on top of the selected virtual lab (for details, see below), a small number of interactive simulations were built to cover the whole range of the required VM affordances.

Fifth, we designed a training intervention for all of our participants, to ensure that they had the knowledge and skills to use the study's PM and VM, as well as, to examine whether our participants had the ability to switch manipulatives (PM to VM and vice versa) during experimentation (for more details on training, see below). For the latter, we also reviewed the relevant literature, in which we found that college students were able to make

transitions from PM experimentation to VM experimentation and vice versa (e.g., Zacharia & Olympiou, 2011), especially when using VM of high fidelity (VM that include concrete representations which resemble reality). In this study, we used VM of high fidelity to ensure smooth transitions between PM and VM. Our VM also carried a tool that provided the corresponding idealized representations of any concrete representation displayed on the computer screen. However, at no point there was a direct transition from PM to the idealized representations of our VM. Concrete representations always intervened between PM and any idealized representations.

Sixth, we revisited the study's experiments and for each one of them we created the PM and VM blend that the students were going to use. The latter requires informing the students which manipulative(s) to use when conducting an experiment. In our case, we adjusted the instructions of each experiment to inform the students when to use PM or VM. Needless to say, the PM and VM blend varied across the study's experiments most of the times. In Appendix A, we provide an example of how objectives of an experiment of the study are outlined and matched with unique PM or VM affordances. Moreover, we provide our reasoning behind each objective–affordance matching. The same process was followed for each one of the study's experiments.

Finally, it should be noted that creating PM and VM blends, according to the afore-mentioned framework, and implementing them in a learning environment requires from researchers or teachers to have certain knowledge and skills. For instance, they need to know which PM and VM are available, how these PM and VM could be used, what affor-dances and limitations PM and VM carry, and whether their students have the knowledge and skills to use them.

THIS STUDY

This study was contextualized through the *Physics by Inquiry* curriculum (McDermott & The Physics Education Group, 1996) aiming to compare the effect of three instructional conditions that differ in the medium (PM or VM) and mode (alone or in combination) of experimentation on undergraduate students' learning in physics, particularly, their under-standing of concepts in the domain of Light & Color. The first condition involved the use of PM (PM condition), the second condition involved the use of VM (VM condition), and the third condition involved the use of a blended combination of PM and VM (PM&VM con-dition) throughout the study. Blending PM and VM was based on the framework described right above.

METHODS

Sample

The participants of the study were 70 (freshmen) undergraduate students of a university in Cyprus (15 males, 55 females; M = 18.3 years, SD = 0.87), who were enrolled in an introductory physics course that was based upon the Physics by Inquiry curriculum (McDermott & The Physics Education Group, 1996). The participants were randomly separated into three conditions, namely, the PM condition (23 students), the VM condition

(23 students), and the PM and VM blended combination condition (PM&VM condition; 24 students). None of the participants had taken college physics prior to the study. The students in all conditions were randomly assigned to groups (three persons in each group) as suggested by the curriculum of the study (McDermott & The Physics Education Group, 1996). However, all the measurements taken in this study targeted the individuals and not their groups as a whole.

Curriculum Materials

The selection of the Physics by Inquiry curriculum was based on the fact that through numerous studies it appeared to enhance undergraduate students' conceptual understanding across physics subject domains (McDermott & Shaffer, 1992; Redish & Steinberg, 1999), including the subject domain of Light & Color. This success of the Physics by Inquiry curriculum is grounded on three foundational components that were found to support conceptual understanding, namely, inquiry, socioconstructivism, and the POE (Predict—Observe–Explain) strategy (see for more details, Zacharia et al, 2008).

For the purposes of this study, three parts of the module of Light & Color were used (McDermott & The Physics Education Group, 1996, p. 225). The first part (Section 1) focuses on an introduction to light, light sources, masks, screens, and shadows, the second part (Section 2) focuses on colored paint, and the third part (Section 3) focuses on colored light. In Section 1, the students are encouraged to develop a mental model that enables them to account for complicated phenomena, such as the formation of images and shadows from extended sources. In Sections 1 and 2, the students conduct experiments with colored paints and colored light, in an attempt to understand how to mix paints of different colors to obtain a particular color of paint and how to combine light of different colors to obtain a particular color of light. Moreover, on the basis of their observations, the students are encouraged to develop a mental model that enables them to predict the color an object will be when viewed under light of different colors.

Finally, it should be noted that in the case of the VM alone and the PM&VM conditions, the wording of this curriculum/teaching material used was slightly modified to refer to the features of the VM.

Material

Physical Manipulatives. PM involved the use of physical instruments (e.g., rulers), objects (e.g., cubes and metal rings), and materials (e.g., lamps, torches, different color filters, projectors) in a conventional physics laboratory. During PM experimentation, feedback was available to the students through the behavior of the actual system (e.g., shape of a shadow on a screen) and through the instruments that were used to monitor the experimental setup (e.g., rulers).

Virtual Manipulatives. VM involved the use of virtual instruments (e.g., rulers), objects (e.g., cubes and metal rings), and materials (e.g., lamps, torches, different color filters, projectors) to conduct the study's experiments on a computer. Most of these experiments were conducted though the virtual laboratory *Optilab*. For a very small number of experiments of the curriculum, the software could not provide all the material needed for the experimental
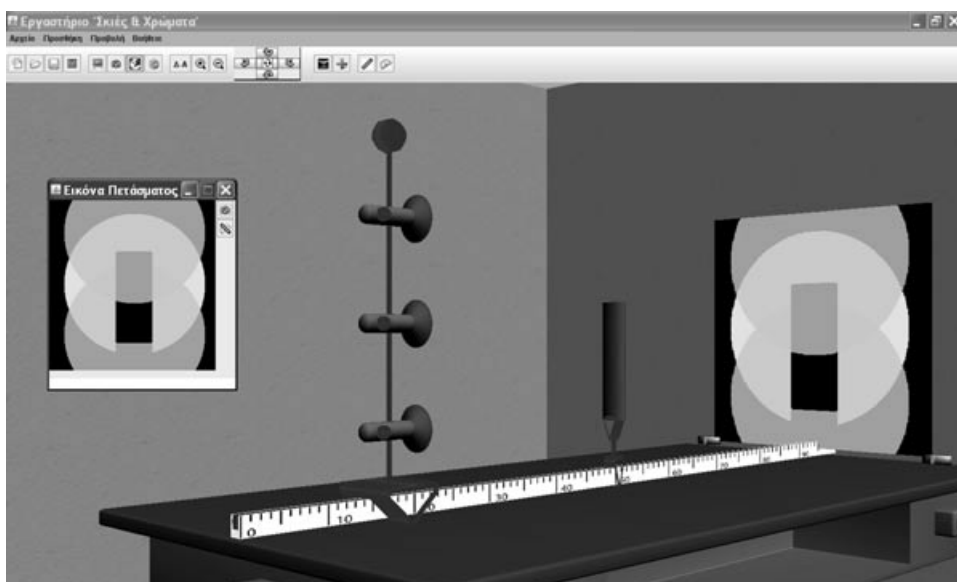
Figure 2. The *Optilab* environment.

setup; hence, interactive simulations were developed and used to complement the Optilab software (Hatzikraniotis, Bisdikian, Barbas, & Psillos, 2007). Optilab (see Figure 2) was selected because of its fidelity and the fact that it retained the features and interactions of the domain of Light & Color as PM did. In its open-ended environment, students of the VM and PM&VM conditions were able to design and conduct the experiments mentioned in the module of Light & Color by employing the "same" material as the ones used by the students using PM.

In the Optilab environment, students were provided with a virtual work-bench on which experiments can be performed, virtual objects to compose the experimental setup, virtual materials whose properties are to be investigated, and virtual instruments (e.g., rulers) or displays (e.g., screen) as illustrated in Figure 2. Students were able to construct their own virtual experimental arrangements by simple and direct manipulation of objects, materials, and virtual instruments. The software offered feedback throughout the conduct of the experiment by presenting information (e.g., distance, color) through the displays of the software. No feedback was provided by the software during the setup of an experiment.

On the one hand, VM could provide analogous feedback to what is routinely available to students through PM. On the other hand, VM carried two additional affordances that PM did not. First, VM offered feedback on the outcome color of any experiment that involved mixing colored paint or combining colored light. PM could not provide such feedback. It was left to the PM students to decide in their groups what the outcome color was. In many cases, the outcome color was a matter of dispute among PM group members, which in some cases proved to be time consuming as well. Second, the VM of the study offered accurate measurements (measurement error-free) at any point of the experiment, which also resulted in saving time and providing to the students opportunities to repeat more easily an experiment than the students in the PM condition did.
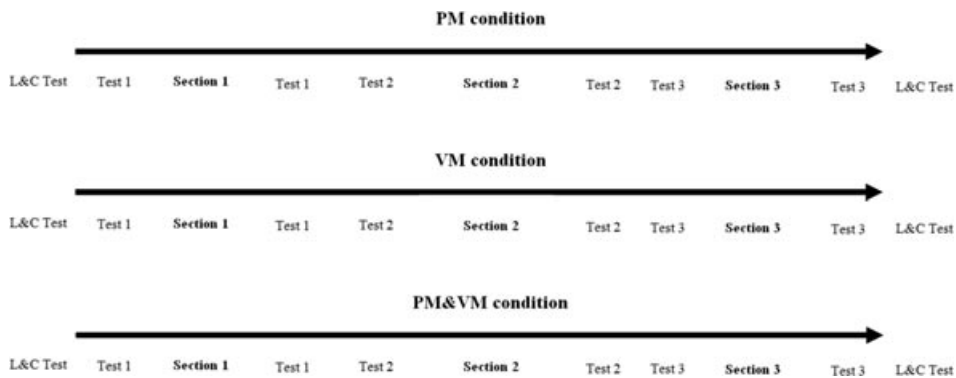
Figure 3. Experimental design of the study.

Data Collection and Treatment

This study involved the collection of data through the use of conceptual tests before, during, and after the study. Specifically, the same conceptual test (Light & Color test or L&C test) was administered to assess students' understanding of Light & Color concepts concerning light, shadows, colored paint, and colored light, both before and after the study. In addition, tests were administered before and after introducing each section (Tests 1, 2 and 3; see Figure 3), with each test being identical before and after each section. The tests were developed and used in previous research studies by the Physics Education Group of the University of Washington (McDermott & The Physics Education Group, 1996). Despite the extensive use of these tests in prior research of the Physics Education Group of the University of Washington (e.g., Wosilait, Heron, Shaffer, & McDermott, 1998), we also piloted, reviewed, and validated (deemed to be appropriate) the content of each test used in this study by a five-member expert panel, namely, two physicists and three physics educators.

Each of the Tests 1, 2, and 3 contained four items (each item consisted of at least two sub-items; for a sample of an item of Test 2, see the first column of the table of the Appendix B) that asked open-ended conceptual questions all of which required explanations of reasoning. The L&C test included five open-ended items assessing all sections of the study's curriculum. This test targeted both the specific concepts introduced in each section, as well as, the interconnections and interdependencies of these concepts. The items that were included in the L&C test were different, but of isomorphic structure and targeting the same concepts, from the items included in the rest of the tests. Each item of each test was scored separately; however, for correct responses a total score was derived from each test and used in the analysis.

All tests were scored and coded blind to the condition in which the student was placed. The scoring of each item was performed through the use of a scoring rubric table that included preset criteria (expected correct answer and expected correct explanation; for an example, see the second, third, and fourth columns of the table of the Appendix B), which were used to score both whether a participant's answer to a question of an item and its accompanied explanation were correct. The maximum score of each question of an item of a test varied according to the number of criteria used for scoring its explanation. Hence, the maximum score of an item of a test varied both across the items of a test and across the items of the other tests, unless two items shared the same total number of explanation criteria. An individual's total score on a test was derived by adding all the assigned scores,

both those of an answer and an explanation, of all questions (of all items) of a test, and by adjusting it to a 100-point scale. The minimum score was 0, and the total maximum score was 100 on each test.

Two independent raters scored about 20% of the data. The reliability measures (Cohen's kappa) for scoring of the L&C test (pre- and posttest) and Tests 1, 2, and 3 (pre- and posttests), were above 0.93 across all tests. The reliability (proportion of agreement) of the scoring of the qualitative data was 0.91. The qualitative analysis concerned students' conceptions as revealed through students' responses to the open-ended items of the conceptual tests. Disagreements were discussed after the reliability analysis and were classified when mutual agreement was reached.

Procedure

First, all conditions were assigned participants after random assignment. Second, within each condition students were randomly assigned to groups (of three) as suggested by the curriculum of the study (in some conditions there was one two-member group because the total number of participants in the group was not enough to form triads). Third, all participants were administered the L&C pretest before getting engaged in the treatment of the condition they belonged to. After a week, (pre-) Test 1 was administered and a brief introduction that aimed to familiarize students with the material they were about to use. All students were introduced to the Physics by Inquiry curriculum and to PM and VM through a demonstration, regardless of their condition. The introduction to the routines and procedures of the Physics by Inquiry curriculum was very important because they differ from those involved in the more traditional, passive modes of instruction that students had experienced during their primary and secondary school years. For example, the enactment of the Physics by Inquiry curriculum does not involve any lecturing, tutoring, or traditional textbook. In contrast, students are seen as responsible for their own learning and are expected to collaboratively construct knowledge and develop their understanding of physics concepts through the conduction of a carefully designed, structured sequence of inquiry-based experiments.

Moreover, the role of the instructors in the Physics by Inquiry curriculum is quite different from that in a traditional instruction. It is supportive in nature and requires instructors' engagement in dialogues with the students of a group at particular points of the activity sequence, as specified by the Physics by Inquiry curriculum. Through these dialogues, the instructors aim to encourage reflection across the inquiry processes and practices involved in the activities of the Physics by Inquiry curriculum and not to lecture or provide readymade answers/solutions. For the purposes of this study, all conditions shared the same five instructors (consisted of one academic and four doctoral students) throughout the instructional intervention. All instructors were previously trained in implementing the Physics by Inquiry curriculum and had experienced its implementation at least for 2 years.

Fourth, along with the instructional part of each section, conceptual tests were also administered both before and after each section (as shown in Figure 3). The duration of the study was 13 weeks. All conditions were facilitated in the same laboratory environment that hosts both conventional equipment and a computer network arranged at the periphery. Students met once a week for one and a half hour. The time-on-task was the same for all conditions.

Data Analysis

The data analysis involved both quantitative and qualitative methods. All tests were scored though the use of scoring rubrics, and the resulted student performance scores were analyzed by using (a) one-way analysis of variance (ANOVA) for the comparison of the pretest scores of the three conditions on each test, (b) paired samples *t*-test for the comparison of the pretest scores to the posttest scores of each condition across all tests, and (c) one-way analysis of covariance (ANCOVA) for the comparison of the posttest scores of the three conditions on each test. For the latter procedure, the students' scores in the corresponding pretests were used as the covariate.

The aim of the first procedure was to determine whether the three conditions of the study were comparable with regard to the sample's entry understanding of physics concepts from the subject domain of Light & Color, before the study and before each section. The aim of the second procedure was to investigate whether the use of the blended combination of PM and VM, and the use of PM, or VM alone, within the context of the Physics by Inquiry curriculum, improved students' conceptual understanding. The aim of the third procedure was to investigate whether the three conditions of the study had differences on the outcome measures (understanding of physics concepts in the domain of Light & Color) of each test. For all analyses, the effect size is also reported (in the case of the ANOVA and ANCOVA, we reported partial η^2; Cohen, 1988). To maintain the overall probability of familywise error (Type I error) at a target level in the case of multiple pairwise comparisons, we applied the Holm's sequentially selective Bonferroni method (Holm, 1979).

The qualitative analysis involved the identification and classification of students' scientifically acceptable conceptions (SACs) and scientifically not-acceptable conceptions (SNACs) concerning *light*, *shadows*, *colored paint,* and *colored light*. The analysis followed the procedures of phenomenography (Marton & Booth, 1997). Phenomenography is used to identify students' qualitatively different, hierarchically related, conceptions of learning. Specifically, for each student's pre- or posttests, the researcher first underlined the most important sentences and marked some keywords that characterized the student's ideas with respect to *light*, *shadows*, *colored paint,* and *colored light*. By comparing the sentences underlined and the keywords derived from the tests, the content-specific similarities and differences between students' test responses about their views on *light*, *shadows*, *colored paint* and, *colored light* were explored and summarized. Then, the researchers constructed "qualitatively different" categories of description, essentially across rather than within the students' responses, that were used to classify the conceptions of *light*, *shadows*, *colored paint,* and *colored light* held by students for each condition separately. By comparing the similarities and differences between the students of each group, some categories for the conceptions of *light*, *shadows*, *colored paint,* and *colored light* emerged. Each category was intended to show a unique way of understanding the phenomenon being researched. Therefore, the purpose of the phenomenographic analysis was to reveal the categories of description that could characterize the qualitatively different perspectives in which *light*, *shadows*, *colored paint* and *colored light* was conceptualized or experienced by the students of each condition. In addition, the prevalence for each one of the resulting categories for each test (L&C pre- and posttests, and Pre- and Posttests 1, 2, 3) was calculated. The aim of the latter was to compare whether the prevalence of students' conceptions changed after the treatments of the study. For the purposes of this paper, a sample of the phenomenographic analysis that concerned conceptions of *colored paint* and *colored light* is included (see the Results section).

TABLE 2
Mean Scores and SD of the PM Condition (PM), the VM Condition (VM), and the PM&VM Condition (PM&VM) in Each of the Tests

Condition	L&C Test	Test 1	Test 2	Test 3
PM pretest	10.5 (4.0)	18.0 (8.0)	1.1 (2.2)	5.4 (5.0)
PM posttest	56.7 (13.6)	49.5 (16.1)	73.7 (22.9)	53.5 (25.1)
VM pretest	13.1 (4.9)	21.4 (8.0)	1.9 (2.7)	8.7 (7.8)
VM posttest	62.6 (14.7)	52.3 (17.7)	83.9 (22.0)	61.4 (24.2)
PM&VM pretest	13.6 (6.8)	23.8 (11.8)	1.5 (2.6)	6.7 (8.6)
PM&VM posttest	75.8 (11.2)	66.6 (16.7)	93.3 (11.0)	76.1 (19.4)

PM condition (PM): participants used PM alone; VM condition (VM): participants used VM alone; PM&VM condition (PM&VM): participants used both PM and VM in a blended mode of experimentation.

RESULTS

Test Performance

The one-way ANOVA procedure indicated that the three conditions did not differ in pretest scores across all of the study's tests ($F < 1$, $p > .05$ across all cases). Means and standard deviations of performance scores are shown in Table 2.

The paired samples t-test showed that all three conditions improved students' understanding of concepts concerning *light and color* both after each section and after the study ($p < .001$ for all comparisons; lower than the 0.004 which is the lowest p-value given by the Holm–Bonferroni method) (see Table 3). However, the ANCOVA procedure revealed differences among the study's three conditions. More specifically, students' scores to the L&C posttest were subjected to an ANCOVA with L&C pretest scores as covariate and condition as between subjects' factor. The analysis revealed a main effect of condition ($F(2, 66) = 10.95$, $p < .001$, partial $\eta^2 = 0.25$) and of L&C pretest scores ($F(1, 66) = 10.42$, $p = .002$, partial $\eta^2 = 0.14$), but no interaction between condition and L&C pretest scores ($F < 1$, $p > .05$).

TABLE 3
The Paired Samples t-test Results for Each of the Study's Tests

Conceptual Tests	Condition	t	df	significance	r
Posttest 1–Pre test 1	PM	10,097	22	<0.0001	.9
Posttest 1–Pre test 1	VM	8,978	22	<0.0001	.88
Posttest 1–Pre test 1	PM&VM	14,045	23	<0.0001	.95
Posttest 2–Pre test 2	PM	15,418	22	<0.0001	.95
Posttest 2–Pre test 2	VM	18,293	22	<0.0001	.96
Posttest 2–Pre test 2	PM&VM	41,565	23	<0.0001	.99
Posttest 3–Pre test 3	PM	9,333	22	<0.0001	.88
Posttest 3–Pre test 3	VM	10,849	22	<0.0001	.91
Posttest 3–Pre test 3	PM&VM	17,728	23	<0.0001	.96
L&C posttest–L&C pre test	PM	17,559	22	<0.0001	.96
L&C posttest–L&C pre test	VM	17,085	22	<0.0001	.96
L&C posttest–L&C pre test	PM&VM	29,831	23	<0.0001	.98

In the case of Test 1, the ANCOVA revealed a main effect of condition (F (2, 66) = 5.104, $p < .01$, partial $\eta^2 = 0.13$) and of Pretest 1 scores (covariate) on students' Posttest 1 scores (F (1, 66) = 14.5, $p < .001$, partial $\eta^2 = 0.18$), but no interaction between condition and Pretest 1 scores ($F < 1$, $p > .05$).

In the case of Test 2, the ANCOVA revealed a main effect of condition (F (3, 66) = 6.07, $p < .01$, partial $\eta^2 = 0.15$). No effect was found either between Pretest 2 scores and students' Posttest 2 scores or between condition and Pretest 2 scores ($F < 1$, $p > .05$ for both cases).

Finally, in the case of Test 3, the ANCOVA also revealed a main effect of condition (F (2, 66) = 5.89, $p < .01$, partial $\eta^2 = 0.15$). No effect was found either between Pretest 3 scores on students' Posttest 3 scores (F (1, 66) = 3.55, $p = .06$, partial $\eta^2 = 0.05$), or between condition and pretest 3 scores ($F < 1$, $p > .05$).

Overall, all of the ANCOVA procedures were found to have a main effect of condition and their effect sizes were found to fall under the small effect category for social science data (0.04–0.25; for details, see Lipsey, 1998 and Ferguson, 2009), with the effect size found in the case of the L&C test reaching the cutoff of the moderate effect category.

Bonferroni-adjusted pairwise comparisons suggested that students' posttest scores in the PM alone and VM alone conditions were significantly lower than those of the students in the PM&VM condition across all tests ($p < .001$ across all PM&VM versus PM or VM comparisons; lower than the 0.016, which is the lowest p-value given by the Holm–Bonferroni method). The pairwise comparisons did not show any significant difference between the students' posttest scores of the PM and VM alone conditions across all tests. These findings suggest that the PM&VM condition enhanced students' understanding of the light and color concepts that were introduced through the curriculum material of this study, more than the PM or VM conditions did. On the other hand, the use of PM or VM alone was equally effective in promoting students' understanding of these light and color concepts.

Understanding of Light and Color Conceptions

The qualitative analysis revealed that the PM alone and VM alone conditions shared mostly the same conceptions across the light and color concepts studied (*light*, *shadows*, *colored paint,* and *colored light*), as either SAC or SNAC, both before and after the L&C test was administered. The same result was found in the analysis of Tests 1, 2, and 3 both before and after the introduction of each section of the study's curriculum. The PM&VM condition appeared to share the same SAC and SNAC conceptions with the other two conditions only before the study at the pretest of each section.

After all students were exposed to the treatments of their conditions, they managed to surpass many of their SNACs and adopt SACs, with the PM&VM condition having the highest increase in percentage across all SACs and the higher decrease in percentage across all SNACs. The latter implies that the blended combination had greater impact on students' transition from SNAC to SAC than the PM and VM alone conditions did (for an example, see Table 4).

DISCUSSION

One of the goals of this study was to develop a framework that portrays how to blend PM and VM for the purposes of laboratory experimentation in science education. Our proposed

TABLE 4
A Sample of SAC and SNAC as They Emerged From the Study's Tests Through the Phenomenographic Analysis

Conceptions (SAC or SNAC[a], Test)	PM Condition[c]		VM Condition[c]		PM&VM Condition[c]	
	Pretests (%) (n)	Posttests (%) (n)	Pretests (%)(n)	Posttests (%) (n)	Pretests (%) (n)	Posttests (%) (n)
Category: How light travels[b]						
Subcategory: Conceptions about when an observer sees an object[b]						
An observer sees an object when light travels from the object to the observer's eye. In some cases, the object generates light. In other cases, the object cannot be seen unless it is illuminated by light from another source. Light from the source reaches the object, reflected from the object and travels from the object to the observer's eye (SAC, Test 1)	26 (6)	82 (19)	17 (4)	70 (16)	16 (4)	100 (24)
An observer sees an object when the object is illuminated by light. It is not a requirement for the light to travel directly to the observer's eye for an object to be seen. (SNAC, Test 1)	74 (17)	18 (4)	79 (18)	30 (7)	84 (20)	0 (0)
An observer sees an object when light travels from the object to the observer's eye. All objects generate light. (SNAC, Test 1)	0 (0)	0 (0)	4 (1)	0 (0)	0 (0)	0 (0)
Category: Color mixing in painting						
Subcategory: Conceptions about mixing colored paint						
Mixing the three primary paint colors (cyan, magenta and yellow) in different ways/combinations and proportions may result in producing all other colors (SAC, Test 2)	0 (0)	70 (16)	0 (0)	70 (16)	0 (0)	96 (23)

(Continued)

TABLE 4
Continued

Conceptions (SAC or SNAC[a], Test)	PM Condition[c]		VM Condition[c]		PM&VM Condition[c]	
	Pretests (%) (n)	Posttests (%) (n)	Pretests (%) (n)	Posttests (%) (n)	Pretests (%) (n)	Posttests (%) (n)
If you mix two colors of paint and then mix the same colors of light, the resulting color will be the same (SNAC, Test 2)	91 (21)	30 (7)	78 (18)	30 (7)	96 (23)	4 (1)
If you mix two colors of paint and then mix the same colors of light, the resulting color will not always be the same (SNAC, Test 2)	9 (2)	0 (0)	22 (5)	0 (0)	4 (1)	0 (0)
Category: Combination of light of different colors						
Subcategory: Conceptions about the white light						
The white light is produced when all the light colors of the spectrum of light are combined before they reach our eyes (SAC, Test 3)	0 (0)	74 (17)	0 (0)	87 (20)	0 (0)	100 (24)
The white light is produced by combining the same colors as in colored paint (SNAC, Test 3)	70 (16)	26 (6)	60 (14)	13 (3)	79 (19)	0 (0)
The white light is a pure color and not a combination of colored light (SNAC, Test 3)	30 (7)	0 (0)	40 (9)	0 (0)	21 (5)	0 (0)

[a]SAC = Scientifically acceptable conception. SNAC = scientifically not acceptable conception.

[b]Thirteen categories of conceptions emerged. Each category of conception consists of several subcategories.

[c]PM condition (PM): participants used PM alone; VM condition (VM): participants used VM alone; PM&VM condition (PM&VM) participants used a blended combination of PM and VM.

framework focuses on the learning objectives of each experiment, which are then associated with the specific manipulatives, PM or VM, that best serve them. This particular association requires a sound understanding of the learning objectives that each experiment aims to fulfil, as well as the affordances/advantages that each manipulative carries, particularly those unique to only PM or VM (Zacharia et al., 2008). Literature in the PM and VM experimentation domain revealed a number of such affordances/advantages that were also

found to be conducive to learning (Hsu & Thomas, 2002; Huppert & Lazarowitz, 2002; Windschitl, 2000). Hence, their presence in learning through experimentation environments becomes vital. For example, only the "messy" interactions with PM teach students about the underlying complexity of doing science (e.g., measurement errors) and give them a more grounded perspective on the limitations of specific virtual environments or, more generally, of VM (Windschitl, 2000); in contrast, VM interactions are the only ones that provide students with opportunities to manipulate reified objects (Zacharia et al., 2008; Triona & Klahr, 2003).

Another goal of this study was to create blends of PM and VM, according to the aforementioned framework, for each one of the experiments included in the study and to compare them to the use of PM and VM alone. The idea was to investigate whether a blended combination of PM and VM is more conducive to students' conceptual understanding than the use of PM or VM alone. As anticipated, our findings revealed that the use of blended combinations of PM and VM across the study's experiments enhanced students' understanding of light and color concepts more than PM or VM alone. A similar pattern was found in the qualitative analysis, in which the students of the PM and VM condition were found to shift from SNAC to SAC to a greater extent across the light and color concepts investigated throughout the study, than the PM or VM alone conditions. These findings indicate that, in the context of this study, the use of a blended combination of PM and VM is more preferable than the use of PM or VM alone.

Furthermore, it is important to highlight the fact that PM and VM were not found to differ across all the tests of the study. However, it should be noted that the mean scores of the students of the VM condition were consistently slightly higher than those of the PM condition across all tests (see Table 2). This small difference in mean scores, in favor of the VM condition, might be attributed to the fact that VM carried more (unique) advantageous affordances than PM. Again, to reach more solid conclusions, further research focusing on the learning process is also necessary, most probably conducted through video data collection and analysis. On the other hand, the fact that these differences were found to be not significant implies that the use of PM and VM, when embedded in a context similar to that of this study, can be equally effective in promoting students' understanding of concepts in the domain of Light & Color. This finding was also confirmed through the qualitative analysis that showed similar shifts from SNAC to SAC in both of the PM and VM alone conditions. This latter finding demonstrates that the nature of learning and the learning outcomes do not substantially change when PM is substituted with VM. This coincides with findings from other research studies which showed that VM can be used (in some contexts and given specific conditions) to provide authentic laboratory experiences that are not substantially different to the methods employed in using PM (Klahr et al., 2007; Triona & Klahr, 2003; Zacharia & Olympiou, 2011).

It should be noted that the present study was carried out in the context of a normal course in physics, thus offering ecological validity to the aforementioned findings. On the other hand, the study had some limitations. The first limitation is that the number of participants was rather small. The second one is that the study targeted only one specific age group, namely undergraduate students, and the third one is that it targeted only one specific topic (Light & Color). Another limitation of this study is that it involved only one data source (conceptual tests). We could have had a better insight in terms of student learning that took place during our intervention, if we had used more data sources, particularly data sources focusing on the process rather than just the end results. For instance, in the case of VM, we could have taken online measures of learning.

Needless to say, more research is needed in this domain before we can reach definite conclusions which can be generalized. There is a need for larger samples of participants, as

well as wider ranges of student ages, subject domains, and PM and VM. Moreover, further research is needed on how to optimize the PM and VM blends. For example, are there any particular PM and VM affordances that should coexist or never be combined? In addition, more frameworks (of the same or different rationale as the one followed in this study) that focus on blending PM and VM affordances to enhance students conceptual understanding should be developed and tested out (in the same and in different subject domain), as well as frameworks that target learning objectives besides conceptual understanding-oriented ones. For instance, how such a framework should look like if the study's learning objectives were focusing on aspects of the nature of science (e.g., what is science and scientific knowledge, distinction between observations and inferences, human error).

Overall, the fact that the use of a blended combination of VM and PM, which is based on a framework similar to the one developed in this study, appears to be more conducive to learning through laboratory experimentation than the use of PM and VM alone, challenges the already established norms concerning experimentation in the science classroom. Specifically, it challenges the laboratory experimentation as we experienced it through PM or VM, in a way that calls for its redefinition and restructuring, to include blended combinations of VM and PM. However, this call for reform creates the need for understanding how PM and VM affordances could affect students' learning. Therefore, it is essential to extend the empirical base through more research focused toward this direction.

APPENDIX A: AN EXAMPLE OF HOW PM AND VM WERE BLENDED ACCORDING TO THE OBJECTIVES OF AN EXPERIMENT
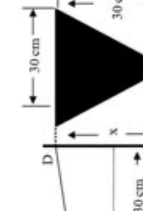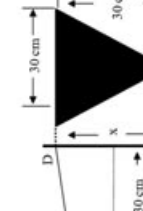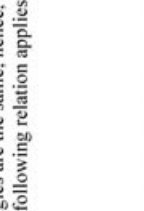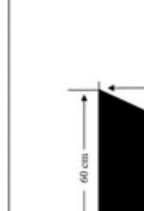
Experiment 6.5[a] (McDermott & The Physics Education Group, 1996, p. 253)

[a]The wording of some experiments of the curriculum was slightly modified to serve the needs of both PM and VM experimentation. The version that the students used indicated specifically which manipulative to use.

APPENDIX B: SCORING RUBRIC FOR ITEM 2 OF TEST 2 (SCORE IN PARENTHESIS)

Question	Correct answer	Expected explanation	Total score
You are given a piece of cardboard in the shape of a triangle. The dimensions of the triangle are shown at right. A. Predict the size and the shape of the shadow that would be formed on the screen if a small lighted bulb, the cardboard triangle and a screen are arranged along a line as shown below. (A line through the bulb and the center of the triangle is perpendicular to the screen.) Explain your reasoning.	The size of the shadow on the screen will be 30 cm (both the width and the length; see the black triangle on the right). The shape and the orientation of the resulted triangle will be the same as the shape of the cardboard triangle. (1)	1. Diagram (1) 2. The right above diagram (sight view) contains 2 triangles: The vertices of the larger triangle (ADE) are located at the bulb and the top and bottom of the bright region on the screen. The vertices of the smaller triangle (ABC) are located at the bulb and at the top and bottom of the cardboard triangle. The corresponding angles of these two triangles are the same; hence, the two triangles are *similar*. Consequently the following relation applies: $60/90 = 10/X$ $X = 900/60$ $X = 15$ cm $\Rightarrow 2 \times 15 = 30$ cm (1)	(3)
B. Is it possible to place the bulb in another location along the dashed line so that the shadow of the triangle is twice as tall as in part A? If so where? If not, why not?	Te bulb should be placed 15 cm away from the cardboard, in order to receive a shadow that it is twice as tall (60 cm) as the size of the shadow in part A. (1)	1. Diagram (1)	(3)

2. The right above diagram (sight view) contains 2 triangles: The vertices of the larger triangle (ADE) are located at the bulb and the top and bottom of the bright region on the screen. The vertices of the smaller triangle (ABC) are located at the bulb and at the top and bottom of the cardboard triangle.

The corresponding angles of these two triangles are the same; hence, the two triangles are *similar*. Consequently the following relation applies:

$$X/X+30 = 20/60$$
$$60X = 20(X+30)$$
$$60X = 20X + 600$$
$$40X = 600 \Rightarrow X = 15 \text{ cm (1)}$$

(3)

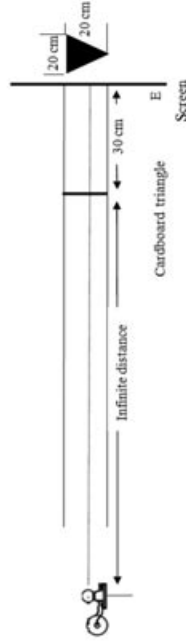|		
C. Is it possible to place the bulb in another location along the dashed line so that the shadow is half as tall as in part A? If so where? If not, why not?	It is not possible to form a shadow half as tall (15 cm) as in part A, no matter where you place the bulb along the dashed line. (1)	1. The construction of such a diagram is impossible. (1) 2. The shadow of the triangle cannot be smaller than the size of the object (cardboard triangle), because of the way the light travels (in the context of the Physics by inquiry curriculum the students learn that the light rays travel in straight lines). Given that light rays travel in line and drawing a light diagram as the one in part B requires having one of the similar triangles vertices located at the bulb, the formation of a triangle that is 15cm tall (smaller than the actual size of the cardboard triangle) is impossible. (1) (3)
D. Suppose that the bulb were placed along the line very far away from the triangle and the screen. What would be the approximate shape and size of the shadow? Explain. Describe the light reaching the screen in this case.	If the bulb was placed at an infinite distance away from the cardboard, the shape and size of the shadow would be approximately the same as the cardboard triangle size and shape. The orientation of the triangle will also remain the same. (1)	1.Diagram (1) 20 cm Infinite distance 30 cm 20 cm Cardboard triangle E Screen 2. If we keep infinitely moving the bulb away from the cardboard, the light rays reaching the cardboard will tend to travel in parallel lines. Thus, in theory, the smallest shadow size someone could get is the actual size of the cardboard triangle. (1) (3)

REFERENCES

Akpan, J. P., & Andre, T. (2000). Using a computer simulation before dissection to help students learn anatomy. Journal of Computers in Mathematics and Science Teaching, 19, 297–313.

Balamuralithara, B., & Woods, P. C. (2009). Virtual laboratories in engineering education: The simulation lab and remote lab. Computer Applications in Engineering Education, 17, 108–118.

Bara, F., Gentaz, E., Pascale, C., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. Cognitive Development, 19, 433–449.

Campbell, J. O., Bourne, J. R., Mosterman, P. J., & Brodersen, A. J. (2002). The effectiveness of learning simulations for electronic laboratories. Journal of Engineering Education, 91, 81–87.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), Mapping the mind. Domain specificity in cognition and culture (pp. 169–200). New York: Cambridge University Press.

Carlsen, D., & Andre, T. (1992). Use of a microcomputer simulation and conceptual change text to overcome student preconceptions about electric circuits. Journal of Computer-Based Instruction, 19, 105–109.

Chen, S. (2010). The view of scientific inquiry conveyed by simulation-based virtual laboratories. Computers & Education, 55, 1123–1130.

Chi, M. T. H., Slotta, J. D., & deLeeuw, N. (1994). From things to processes a theory of conceptual change for learning science concepts. Learning and Instruction, 4, 27–43.

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

deJong, T. (2006). Computer simulations: Technological advances in inquiry learning. Science, 312, 532–533.

deJong, T., & Njoo, M. (1992). Learning and instruction with computer simulation: Learning processes involved. In E. de Corte, M.C. Linn, H. Mandl, & L. Verschaffel (Eds.), Computer-based learning environments and problem solving (pp. 411–427). Berlin: Springer-Verlag.

deJong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. Review of Educational Research, 68, 179–202.

diSessa, A. A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy (from the "pieces" side of the fence). In S. Vosniadou (Ed.), International handbook of research on conceptual change (pp. 35–60). New York: Routledge.

Doerr, H. (1997). Experiment, simulation and analysis: An integrated instructional approach to the concept of force. International Journal of Science Education, 19, 265–282.

Engle, R. A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. Journal of the Learning Sciences, 15, 451–498.

Faryniarz, J. V., & Lockwood, L. G. (1992). Effectiveness of microcomputer simulations in stimulating environmental problem solving by community college students. Journal of Research in Science Teaching, 29, 453–470.

Feisel, L. D., & Rosa, A. J. (2005). The role of the laboratory in undergraduate engineering education. Journal of Engineering Education, 94, 121–130.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research and Practice, 40, 532–538.

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., et al. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. Physical Review Special Topics—Physics Education Research, 1, 1–8.

Gire, E., Carmichael, A., Chini, J. J., Rouinfar, A., Rebello, S., Smith, G., et al. (2010). The effects of physical and virtual manipulatives on students' conceptual learning about pulleys. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), Learning in the disciplines: Proceedings of the 9th international conference of the learning sciences (ICLS 2010) (Vol. 1, pp. 937–944). Chicago: International Society of the Learning Sciences.

Girod, M., Rau, C., & Schepige, A. (2003). Appreciating the beauty of science ideas: Teaching for aesthetic understanding. Science Education, 87, 574–587.

Girod, M., & Wong, D. (2002). An aesthetic (Deweyan) perspective on science learning: Case studies of three fourth graders. Elementary School Journal, 102, 199–224.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. Journal of the Learning Sciences, 14, 69–110.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. Reading Research Quarterly, 28, 117–155.

Hatzikraniotis, E., Bisdikian, G., Barbas, A., & Psillos, D. (2007). Optilab: Design and development of an integrated virtual laboratory for teaching optics. In C. P. Constantinou, Z. C. Zacharia, & M. Papaevripidou

(Eds.), Proceedings of the 7th International Conference on Computer Based Learning in Science. Crete, Greece: Technological Educational Institute of Crete.

Hofstein, A., & Lunetta, V. (2004). The laboratory in science education: Foundations for the twenty-first century. Science Education, 88, 28–54.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

Hsu, Y.-S. (2008). Learning about seasons in a technologically enhanced environment: The impact of teacher-guided and student-centered instructional approaches on the process of students' conceptual change. Science Education, 92, 320–344.

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

Huppert, J., & Lazarowitz, R. (2002). Computer simulations in the high school: Students' cognitive stages, science process skills and academic achievement in microbiology. International Journal of Science Education, 24, 803–821.

Jaakkola, T., & Nurmi, S. (2008). Fostering elementary school students' understanding of simple electricity by combining simulation and laboratory activities. Journal of Computer Assisted Learning, 24(4), 271–283.

Jaakkola, T., Nurmi, S., & Veermans, K. (2010). A comparison of students' conceptual understanding of electric circuits in simulation only and simulation-laboratory contexts. Journal of Research in Science Teaching, 48, 71–93.

Jonassen, D. H. (2000). Integrating constructivism and learning technologies. In J. M. Spector & T. M. Anderson (Eds.), Integrated and holistic perspectives on learning, instruction and technology. Boston: Kluwer Academic Publishers.

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. Journal of Research in Science Teaching, 44, 183–203.

Klatzky, R. L., & Lederman, S. J. (2002). Touch. In A. F. Healy & R. W. Proctor (Eds.), Experimental psychology (pp. 147–176). New York: Wiley.

Kim, M. C., Hannafin, M. J., & Bryan, L. A. (2007). Technology-enhanced inquiry tools in science education: An emerging pedagogical framework for classroom practice. Science Education, 91, 1010–1030.

Limon M., & Mason, L. (2002). Reconsidering conceptual change: Issues in theory and practice. Dordrecht, The Netherlands: Kluwer.

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), Handbook of applied social research methods (pp. 39–68). Thousand Oaks, CA: Sage.

Loomis, J. M., & Lederman, S. J. (1986). Tactual perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performances, Vol. 2, Cognitive processes and performance (pp. 31/1–31/41), New York: Wiley.

Maisch, C., Ney, M., van Joolingen, W. R., & de Jong, T. (2009). Inconsistent reasoning about measurement by undergraduate physics students. Paper presented at the ESERA International Conference of the European Science Education Research Association, Istanbul (Turkey).

Marshall, J. A., & Young, E. S. (2006). Preservice teachers' theory development in physical and simulated environments. Journal of Research in Science Teaching, 43, 907–937.

Martínez-Jiménez, P., Pones-Pedrajas, A., Climent-Bellido, M. S., & Polo, J. J. (2003). Learning in chemistry with virtual laboratories. Journal of Chemical Education, 80, 346–352.

Marton, F., & Booth, S. (1997). Learning and awareness. Mahwah, NJ: Erlbaum.

McDermott, L. C., & Shaffer, P. S. (1992). Research as a guide for curriculum development: An example from introductory electricity, Part I: Investigation of student understanding. American Journal of Physics, 60, 994–1003.

McDermott, L. C., & The Physics Education Group. (1996). Physics by inquiry. New York: Wiley.

Mosterman, P. J., Dorlandt, M. A. M., Campbell, J. O., Burow, C., Bouw, R., Brodersen, A. J., & Bourne, J. R. (1994). Virtual engineering laboratories: Design and experiments. Journal of Engineering Education, 83, 279–285.

National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.

Piaget, J. (1985). The equilibration of cognitive structure. Chicago: University of Chicago Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of scientific conception: Toward a theory of conceptual change. Science Education, 66, 211–227.

Redish, E. F., & Steinberg, R. N. (1999). Teaching physics: Figuring out what works. Physics Today, 52, 24–30.

Ronen, M., & Eliahu, M. (2000). Simulation a bridge between theory and reality: The case of electric circuits. Journal of Computer Assisted Learning, 16, 14–26.

Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. Educational Researcher, 20, 2–9.

Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. Cognition and Instruction, 22, 393–429.

Tao, P., & Gunstone, R. (1999). The process of conceptual change in force and motion during computer-supported physics instruction. Journal of Research in Science Teaching, 36, 859–882.

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. Cognition and Instruction, 18, 423–459.

Toth, E. E., Morrow, B. L., & Ludvico, L. R. (2009). Designing blended inquiry learning in a laboratory context: A study of incorporating hands-on and virtual laboratories. Innovative Higher Education, 33(5), 333–344.

Triona, L., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. Cognition and Instruction, 21, 149–173.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), International handbook of research on conceptual change (pp. 3–34). New York: Routledge.

Windschitl, M. (2000). Supporting the development of science inquiry skills with special classes of software. Educational Technology Research and Development, 48, 81–95.

Winn, W., Stahr, F., Sarason, C., Fruland, R., Oppenheimer, P., & Lee, Y.-L. (2006). Learning oceanography from a computer simulation compared with direct experience at sea. Journal of Research in Science Teaching, 43, 25–42.

Wosilait, K., Heron, R., L. P., Shaffer, S. P., & McDermott, L. C. (1998). Development and assessment of a research-based tutorial on light and shadow. American Journal of Physics, 66(10), 906–913.

Wu, H.- K., Krajcik, J. S., & Soloway, E. (2001). Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom. Journal of Research in Science Teaching, 38, 821–842.

Yueh, H. P., & Sheen, H. J. (2009). Developing experiential learning with a cohort-blended laboratory training in nano-bio engineering education. International Journal of Engineering Education, 25, 712–722.

Zacharia, Z. C. (2005). The impact of interactive computer simulations on the nature and quality of postgraduate science teachers' explanations in physics. International Journal of Science Education, 27, 1741–1767.

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

Zacharia, Z. C., & Anderson, O. R. (2003). The effects of an interactive computer-based simulations prior to performing a laboratory inquiry-based experiments on students' conceptual understanding of physics. American Journal of Physics, 71, 618–629.

Zacharia, Z. C., & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the physics by inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. American Journal of Physics, 76, 425–430.

Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. Learning & Instruction, 21, 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, 1021–1035.

Zollman, D. A., Rebello, N. S., & Hogg, K. (2002). Quantum mechanics for everyone: Hands-on activities integrated with technology. American Journal of Physics, 70, 252.