The Effects on Students’ Conceptual Understanding of Electric Circuits of Introducing Virtual Manipulatives Within a Physical Manipulatives-Oriented Curriculum

Zacharias C. Zacharia & Ton de Jong

To cite this article: Zacharias C. Zacharia & Ton de Jong (2014) The Effects on Students’ Conceptual Understanding of Electric Circuits of Introducing Virtual Manipulatives Within a Physical Manipulatives-Oriented Curriculum, Cognition and Instruction, 32:2, 101-158, DOI: 10.1080/07370008.2014.887083

To link to this article: http://dx.doi.org/10.1080/07370008.2014.887083

Published online: 01 Apr 2014.
The Effects on Students’ Conceptual Understanding of Electric Circuits of Introducing Virtual Manipulatives Within a Physical Manipulatives-Oriented Curriculum

Zacharias C. Zacharia  
*University of Cyprus*

Ton de Jong  
*University of Twente*

This study investigates whether Virtual Manipulatives (VM) within a Physical Manipulatives (PM)-oriented curriculum affect conceptual understanding of electric circuits and related experimentation processes. A pre–post comparison study randomly assigned 194 undergraduates in an introductory physics course to one of five conditions: three experimental conditions with different PM and VM sequences and two control conditions with only PM or VM. Conceptual tests assessed students’ understanding. Instructors’ journals, video data, and interviews provided process-related data. Results showed interplay between manipulative and circuit types. For simple circuits, PM and VM use similarly impacted students’ understanding. However, VM better facilitated understanding than PM for complex circuits: PM users, unlike VM users, encountered process-related problems that prevented development of an appropriate conceptual model because only VM afforded a view of current-flow. When students used VM before PM for complex circuits, they developed the appropriate conceptual model to use in the PM phase.

In engineering and science education, physical laboratory work (also called physical or real laboratories or hands-on activities) is traditionally used to provide students with an additional and more experiential form of learning than through textbook or lecture. Modern technology now offers an alternative for physical laboratory work in the form of virtual computer-based laboratories (Waldrop, 2013). Virtual laboratories have the obvious advantages of being cheaper to use (once they are set up) and in some cases safer than physical laboratories, and they are readily accessible at any time. Evidence is also now accumulating that virtual laboratories compare favorably with physical laboratories in their effectiveness for supporting the acquisition of conceptual knowledge. However, there are also indications that combinations of physical and virtual laboratories may provide students with the optimal affordances for acquiring conceptual knowledge. In the current study we aimed to understand the effects on students’ conceptual
understanding of electric circuits after introducing virtual laboratories, which we will call virtual manipulatives (VM) in our study, within an existing inquiry curriculum that is based on the use of physical manipulatives (PM). We investigated the effects of learning with different sequences of PM and VM use over the course of the study, as well as with PM or VM exclusively. The idea behind the use of different sequences was to investigate whether the introduction of VM at different points in the curriculum would have different effects on the students’ experimentation processes and conceptual understanding. In this study, VM and PM had some similarities in their affordances for experimentation (e.g., setting up electric circuits), as well as some differences (e.g., PM provided touch sensory input, and VM provided a reified abstract or conceptual object, namely, a view of the charge flow). Before presenting our study, we discuss the presumed advantages of physical, virtual, and combined laboratories and summarize existing research.

**ADVANTAGES OF PHYSICAL AND VIRTUAL LABORATORIES**

Laboratories are used in education for a multitude of reasons. Balamuralithara and Woods (2009) list 13 educational objectives related to the use of laboratories. Some of these, including the acquisition and development of psychomotor skills, awareness of following and implementing safety procedures, and learning how to use human sensory input for collecting data, are more prominently associated with physical laboratories (and are often regarded as involved in doing authentic science). However, a central educational objective, the acquisition of conceptual knowledge by the testing of theoretical models, experimentation, and data analysis, is also associated with the use of virtual, simulated laboratories (de Jong, 2006a). Therefore, it makes sense to compare the (presumed) advantages of both approaches with regard to the acquisition of conceptual knowledge (see also de Jong, Linn, & Zacharia, 2013).

One possible advantage of physical laboratories for acquiring conceptual knowledge relates to the role of touch sensory input (Zacharia, Loizou, & Papaevripidou, 2012). It has been argued that touch sensory input may reduce the cognitive load of a learner’s working memory and thus support more complex understandings (Zacharia & Olympiou, 2011). In this framework, each modality (visual, auditory, touch) has its own processing channel (Burton & Sinclair, 2000; Millar, 1999). Cognitive load could be reduced either because touch sensory input can transfer the same information with the sensory input transferred through the visual or the auditory channels (same information is distributed to more than one processing channels), or because the touch sensory channel transfers different (but complementary) information to the sensory input transferred through the visual or auditory channels (for details see Zacharia & Olympiou, 2011). Arguments for the advantages associated with touch sensory input are given for learning in different domains such as biology (Jones, Andre, Superfine, & Taylor, 2003; Jones, Minogue, Tretter, Negishi, & Taylor, 2006), physics (Enyedy, Danish, Delacruz, & Kumar, 2012), chemistry (Bivall, Ainsworth, & Tibell, 2011), and even abstract domains such as mathematics (Carlson, Avraamides, Cary, & Strasberg, 2007; Clements, 1999). This alleged advantage pertaining to physical laboratories can now, in some cases, also be afforded by simulated laboratories, even in cases where no real world physical counterpart exists. For example, Jones et al. (2003, 2006) added a haptic component to a simulation of viruses at nanoscale. They found that adding a haptic component increased students’ motivation. In a similar vein, Bivall et al. (2011) added a haptic component to a simulation such that students could “feel” interactions between molecules as forces. This
helped students to acquire better conceptual knowledge than using the same simulation without the haptic component, and students’ explanations of their answers at a test were more force-based.

One specific advantage for virtual laboratories that may support the acquisition of conceptual knowledge is that reality can be adapted to serve the learning process. Reality can be simplified by taking out details (and thus lowering fidelity, see e.g., Bell & Trundle, 2008) or by changing model characteristics such as the timescale (see e.g., Ford & McCormack, 2000). It can also be augmented by adding specific features to reality such as the flow of electric current in an electric circuit (Jaakkola, Nurmi, & Lehtinen, 2010), electric and magnetic fields (Ibáñez, Di Serio, Villarán, & Delgado Kloos, 2014), momentum or kinetic energy (Marshall & Young, 2006), or light rays (Olympiou, Zacharias, & de Jong, 2013). Reducing fidelity means that the requirements on students are less burdensome, which may aid learning (Alessi, 1988). Augmenting reality means that concepts that are not visible for students in physical laboratories now become visible. This could lead students to spend more time on processing the feedback in the virtual laboratory as compared to giving more attention to performing procedural actions in a physical laboratory (Marshall & Young, 2006).

It is a general finding in the literature that inquiry learning is only effective if students receive sufficient instructional guidance (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Furtak, Seidel, Iverson, & Briggs, 2012). Instructional guidance can come in many different forms, ranging from simply showing learners their activities in the inquiry task, through scaffolding, to direct prompts (de Jong & Lazonder, 2014). Teachers or lab assistants can give guidance in physical laboratories, but in virtual labs this guidance can be offered to each individual student just in time and in a personalized way (see, e.g., de Jong, 2006b; de Jong et al., 2013; Quintana et al., 2004).

When students engage in inquiry learning in a laboratory, both physical and virtual laboratories (use of PM or VM) have specific advantages as far as the ease of experimentation. Whereas virtual laboratories offer quick access to experimentation, in physical laboratories extensive time is often required to set up equipment, and experiments cannot be redone quickly. Even though this may seem to be a disadvantage, it also means that there is a natural tendency for students to reflect before acting. With virtual laboratories, students can experiment without any costs and thus may end up conducting random, non-goal-directed trials (de Jong & van Joolingen, 1998) or poorly designed experiments (Renken & Nunez, 2013). With physical laboratories, students need to consider costs and reflect carefully on their ideas before engaging in experimentation. On the other hand, it could also be an advantage of virtual laboratories that students have options for easy experimentation, so that ideas can be quickly tested and evaluated (Carlsen & Andre, 1992; Huppert, Lomask, & Lazarowitz, 2002; Zacharia, Olympiou, & Papaevripidou, 2008).

WHAT DOES RESEARCH SAY ABOUT THE EFFECTIVENESS OF THE USE OF VIRTUAL AND PHYSICAL LABORATORIES?

Direct comparisons of the use of virtual and physical experimentation cover a range of domains and types of students and often measure conceptual knowledge as well as assess the effect on experimentation skills. Overall, results show that virtual laboratories seem to be as effective or more effective than physical ones for acquiring conceptual knowledge (de Jong et al., 2013). There is a series of studies in which no differences were found between students’ performance when using either a virtual or a physical lab, covering a wide range of domains, including chemistry...
(Tatli & Ayas, 2013; Wiesner & Lan, 2004), car design (Klahr, Triona, & Williams, 2007), heat and temperature (Zacharia & Constantinou, 2008), polyominoes (Yuan, Lee, & Wang, 2010), oceanography (Winn et al., 2006), spring behavior (Siler, Mowery, Magaro, Willows, & Klahr, 2010; Triona & Klahr, 2003), and pulleys (Renken & Nunez, 2013). However, other work shows benefits from the use of virtual laboratories compared to the use of physical laboratories that can be related to the supposed advantages of virtual laboratories (Zacharia, 2007; Zacharia et al., 2008).

The most frequently mentioned explanation for better performance in a virtual laboratory is found in the ability to provide an adaptation of reality. Huppert et al. (2002), for example, compared two groups of microbiology students; one group followed a traditional classroom approach with an additional physical laboratory, and the other group learned with a computer simulation partially replacing the laboratory. They found the simulation group had better scores on a conceptual test and also some advantages in acquiring science process skills. According to Huppert et al. (2002), the explanation for these findings is that the virtual laboratory made the process simpler for the students; for example, by presenting results directly in visuals, students could decide whether they could go onto the next step in the process. Finkelstein et al. (2005) found that students working with simulated electric circuits outperformed their counterparts in a physical laboratory on quality and efficiency in building circuits, on explaining the working of circuits, and on a conceptual test on electric circuits. They mainly attributed this to the fact that in the virtual laboratory, students could observe the charge flow. This phenomenon was confirmed in a slightly different setup by Olympiou et al. (2013), who showed that adding abstract concepts (light rays) to a concrete virtual optics laboratory (light and color) helped students to better understand the domain. Like Huppert et al. (2002), Finkelstein et al. (2005) also mention that their simulation limited the choice of (irrelevant) variables to manipulate (e.g., the color of the wires), which made the design of experiments simpler for learners. Pyatt and Sims (2012) compared students working with a virtual or a physical environment in a chemistry domain, and found that in one case the physical and virtual environment led to similar outcomes, but that in another case students in the virtual lab tended to create less structured experiments due to the greater ease of experimentation. Other work has found advantages of virtual laboratories over physical ones and attributed this difference to clearer observations, simplified environments, and augmented reality (Martinez, Naranjo, Perez, Suero, & Pardo, 2011). Students also recognize the advantages of virtual labs; when students responded to a survey on learning in virtual and physical laboratories, they indicated that virtual laboratories better enable them to be aware of the experimental design and that virtual laboratories provide space for critical thinking (Koretsky, Kelly, & Gummer, 2011).

Two types of results have been found concerning ease of experimentation. First, Huppert et al. (2002) claim that students learned better in their virtual laboratory because they could perform more experiments and were able to repeat experiments with different controlling variables more easily. Second, Renken and Nunez (2013), who did not find a difference in learning between a virtual and physical laboratory, note that students in the virtual lab tended to create less structured experiments due to the greater ease of experimentation.

The role of instructional support was emphasized by Chang, Chen, Lin, and Sung (2008), who compared students who worked with a physical optics laboratory with students learning with versions of a virtual lab that had different types of instructional support. They found that students
working with one of the virtual laboratories scored better on a test of conceptual knowledge than students in the physical laboratory, and they attributed that to the support that was offered in the virtual environments.

The studies above indicate overall that students achieved identical levels of performance in learning from virtual and physical laboratories or showed an advantage in virtual laboratories. One recent study by Zacharia et al. (2012) does not fit into this pattern. These authors explored the role of touch sensory input with very young children and found that children with wrong conceptions about a balance beam profited more from learning with tangible or physical materials. Moreover, they conjectured that touch sensory input might not be necessary for learners (not necessarily young ones) that have already established the tactual sensory-related knowledge needed for understanding a phenomenon or system (from previous experience).

Overall, it seems that there are a number of good reasons to prefer virtual over physical laboratories. Virtual environments are as good as or better than physical ones for supporting learning. These results have been found for many domains, in different contexts, and for different age levels (with the exception of very young children) and are commonly attributed to the affordances that VM provide to the learner. These affordances include simplifying the real world (e.g., minimization or exclusion of error, focusing only on certain aspects of the phenomenon, and avoiding the messy nature of science); proving representations for unobservable and abstract concepts, variables or phenomena; providing multiple depictions of results; making the setup, modifications, and repetition of experiments easy; and providing integrated guidance or support (de Jong et al., 2013). However, there are also indications in the literature that combinations of both approaches may provide students with the optimal learning opportunities, as will be reviewed in the next section.

USING BOTH VIRTUAL AND PHYSICAL LABORATORIES

Traditionally there has been a separation between virtual and physical laboratories, but recently there has been work that has started to develop and investigate combinations and sequences of the two. There are different possibilities here: blending (e.g., Yueh & Sheen, 2009) possibly with the use of remote laboratories (Jara, Candelas, Puente, & Torres, 2011; van Jooolingen, de Jong, Lazonder, Savelsbergh, & Manlove, 2005), and alternating both modes for the same (Campbell, Bourne, Mosterman, & Brodersen, 2002; Jaakkola & Nurmi, 2008; Kollöffel & de Jong, 2013) or different content (e.g., Zacharia, 2007). We could not find studies on blended environments that investigated effects on the acquisition of conceptual knowledge, but there are such studies on alternating environments for the same and different content.

Climent-Bellido, Martínez-Jiménez, Pones-Pedrajas, and Polo (2003) compared knowledge acquisition of students working with only a physical laboratory on the chemistry topic of distillation with students who had a simulation preceding the physical laboratory. Their results showed that students who used both environments reached higher levels of conceptual knowledge. Zacharia (2007) had two groups of students explore different aspects of electric circuits over an entire curriculum. One group followed the curriculum using only a physical laboratory, while the other group started off with a virtual laboratory and moved to a physical laboratory halfway through the course. The group that switched laboratories scored better than the group that used only the physical laboratory on a conceptual posttest (administered after both groups had
completed the final physical laboratory. The alternating group already had an edge in conceptual knowledge halfway through the course, which indicates that the use of the virtual environment better promoted the acquisition of conceptual knowledge. This advantage was obviously maintained during the portion in which both groups used the physical laboratory. Zacharia et al. (2008) tested the opposite order of combining the two types of environments. They compared a group of students who learned with a physical laboratory (on heat and temperature) with a group of students who first used a physical laboratory and then moved to a virtual one. Results on a test of conceptual knowledge showed no differences between the two groups after the first half of the curriculum (both groups using the physical laboratory); at the end of the course, the students of the group that switched to the virtual laboratory scored better. In earlier work by Akpan and Andre (2000) investigating students who were learning about dissecting a frog, it was found that students who worked with a simulation prior to actual dissection achieved higher scores on a test measuring knowledge of frog anatomy than students who worked only with the physical laboratory or who worked with the physical laboratory prior to using a simulation. However, students using only the simulation environment scored at the same level as students having the simulation precede the physical laboratory.

In these studies, the better performance seen in the alternation condition may still be due to the effect of the virtual environment (simulation) per se. Jaakkola and Nurmi (2008), however, had students complete assignments on electric circuits under three conditions: a virtual lab, a physical lab, and a condition in which students first worked through the virtual lab and then did the same assignments in the physical lab. They found the alternating condition to be the most advantageous for acquiring conceptual knowledge, followed by the simulation environment, with the physical laboratory condition yielding the lowest scores. In another work by Jaakkola, Nurmi, and Veermans (2011), again in the field of electric circuits, students first constructed a circuit by means of a simulation and then by using physical equipment. Their results showed that using both environments was more favorable for students’ resulting quality of knowledge than having students work only in a simulation environment.

Other research studies in this domain have used a balanced design in terms of the order of using PM and VM. For example, Chini, Madsen, Gire, Rebello, and Puntambekar (2012) compared a virtual–physical and a physical–virtual sequence in order to investigate if students’ learning in the domain of pulleys would vary and found no differences. On the other hand, other work showed an advantage of learning first in the virtual laboratory. Following the same crossing of conditions (virtual–physical versus physical–virtual), Toth, Morrow, and Ludvico (2009) reported a small but nonsignificant advantage for the virtual–physical sequence on a conceptual knowledge test in the domain of DNA gel electrophoresis. In a follow-up study in the same domain by Toth, Ludvico, and Morrow (2012), the difference in favor of the virtual-laboratory-first group was more prominent. Students who followed the virtual–physical sequence outperformed students who followed the reverse order of laboratories on tests of experiment design and conceptual knowledge.

The overall conclusion from the literature seems to be that virtual laboratories should, generally speaking, be preferred over physical ones for the development of conceptual knowledge. Touch sensory input, as present in physical laboratories, does not improve learning as compared to virtual laboratories, with the possible exception of learning for individuals, especially very young children, with little prior knowledge (de Jong et al., 2013). The advantage of virtual laboratories is attributed to such features as showing aspects of the domain that are not visible in real laboratories.
Another trend seen in the literature is that benefits can be gained from some type of alternation of physical and virtual laboratories. It is not yet quite clear why this may be the case. Jaakkola et al. (2010) report a study in which they videotaped and compared students who constructed electric circuits only in a simulated environment with students who first did this virtual construction and then made the same circuit with physical material. The video data analysis showed that the students in the virtual and physical laboratory condition benefited from the fact that they had to compare the observations coming out of the virtual and physical laboratories. These comparisons were especially helpful when these observations differed, and students had to go into abstract reasoning to elucidate these differences. Toth et al. (2012) assert that the advantages of the virtual–physical sequence can be attributed to the facts that through simplification and augmentation, students successfully acquire the basic concepts in the virtual environment and that they can then take that knowledge to the more complex and often more blurry physical laboratory. This would mean that if the two types of laboratories cover different subtopics, students could still profit from experience in a virtual laboratory before a physical one. Advantages in alternating different types of laboratories across different subtopics throughout a curriculum were found in Zacharia (2007). In the current study, we evaluated the effects of different sequences of using virtual and physical laboratories and looked further at students’ experimentation processes and activities associated with the different sequences.

**THIS STUDY**

The aim of the present study was to investigate whether introducing virtual laboratories (which we will refer to as VM) within an existing inquiry curriculum that is geared toward the use of physical laboratories (which we will refer to as PM), has a differential effect on students’ conceptual understanding of electric circuits, and whether any possible differences relate to the processes in which students engage during PM or VM experimentation. In this study, VM and PM had some similar affordances for experimentation (e.g., setting up electric circuits), as well as some differences (e.g., PM provided touch sensory input, and VM provided access to an abstract or conceptual object, namely, a view of the charge flow that shows dots moving at various speeds). We selected the subject domain of electric circuits because it was found in previous research studies in this domain that complementing PM with VM carries differing unique affordances and is more conducive to students’ learning than using PM alone (Jaakkola & Nurmi, 2008; Zacharia, 2007). Given the aforementioned literature review on the VM unique affordances, we conjectured that VM would be more beneficial to students’ learning than PM in the context of electric circuits when reality must be augmented to provide students with a view of the current flow (Finkelstein et al., 2005; Jaakkola et al., 2010). Students must be granted quick access to experimentation (easy experimentation, so that ideas can be quickly tested and evaluated) without being distracted by the “messy” nature of “authentic” science (as provided through PM) (Zacharia, 2007).

The selection of the subject domain was a critical aspect for our research design. In order to be able to examine the differential effect on students’ conceptual understanding of electric circuits when implementing VM at different points in a PM-dominant curriculum, we needed to use VM that have been shown to positively influence students’ learning in this particular subject.
domain. In the case of electric circuits, there is strong evidence that the use of VM could enable students to overcome both conceptual and process-related problems (e.g., continuous access to observations) that have been found to limit student learning about electric circuits when only PM are used (Finkelstein et al., 2005; Jaakkola & Nurmi, 2008; Zacharia, 2007).

There are quite a few curricula that involve goals (over and above the promotion of conceptual understanding) that are PM dependent, such as understanding the messy nature of science, developing psychomotor skills related to the use of material and equipment in an experimental setup, or identifying and handling measurement error. Therefore, it is interesting to examine how a VM environment with an additional and advantageous affordance (i.e., view of charge flow) could fit into a context that requires PM presence. In doing so, we need to investigate the circumstances under which the implementation of VM could optimize students’ conceptual understanding in a PM-dominant context. For instance, is it possible to enhance students’ conceptual understanding by using abstract or conceptual objects (i.e., view of charge flow) only for specific parts of the curriculum in the same way as when using such objects across all parts of the curriculum? If so, which parts of the curriculum are they? Do they relate to any particular problems that students encounter (e.g., conceptual, process-related)? These are important questions that have been left unanswered so far, and, in our perspective, answering them is crucial for understanding which are the critical points in a curriculum for combining PM and VM. In order to reach such conclusions, we introduced VM at different parts of the curriculum for different topics and did an in-depth analysis of the associated student experimentation behaviors (the processes in which students engaged during experimentation) and their resulting conceptual understanding.

In the context of this study, experimentation refers to a methodical procedure (including setting up the experiment) carried out with the goal of verifying, falsifying, or establishing the validity of a prediction or a hypothesis. Experimentation processes refer to (a) stating a hypothesis or a prediction, (b) selecting proper materials and using them to set up the experiment, (c) running the experiment (a process of manipulating variables in an effort to determine how they affect the outcome), (d) making observations and determining their validity (the latter depends on whether the students accept the circuit built to be valid, that is, to be the circuit that the experiment required to be built), and (e) reaching conclusions (e.g., whether the observations or outcomes match the predictions or hypotheses). Given that students were required to work in groups of three throughout the curriculum, these processes were all enacted collaboratively, which means that each process involved discussion among group members.

METHOD

Design

For the purposes of this study we used the electric circuits module in *Physics by Inquiry* curriculum (McDermott & the Physics Education Group, 1996, p. 382–454). PM and VM were implemented within a three-part curriculum unit (for more information, see the Curriculum Materials section). A pretest–posttest experimental design was used, as shown in Figure 1. A high fidelity VM environment was used; it retained the features and interactions of the subject domain of the study just as PM did and carried a unique affordance that had been found in prior research to
positively impact students’ learning (e.g., Finkelstein et al., 2005). Specifically, the VM provided a representation of abstract or conceptual objects (i.e., charge flow).

The experimental study carried out involved the comparison of five conditions: a condition that included the use of PM throughout the study (PPP condition), a condition that included the use of VM throughout the study (VVV condition), and three sequential alternations of PM and VM, namely the VPP, PVP and PPV conditions, in which VM was used for different parts of the curriculum.

The VVV and PPP (control) conditions were included because we wanted to establish baselines in each of the parts of the curriculum material and to ensure that alternating between PM and VM did not encumber students’ learning through experimentation in any way. We also wanted to make sure that learning though either using PM or VM alone in the context of this study was feasible (e.g., to verify that PM and VM each did provide useful supports for learning). The rationale behind the configuration of the VPP, PVP, and PPV conditions was to understand the effect that using the study’s VM has on the learning of each of these parts. The VPP, PVP, and PPV conditions were also used to determine if whether knowledge gained during VM use is transferred to the subsequent parts of the study’s curriculum, particularly in the cases of the VPP and PVP conditions. For the PPV condition, our interest focused on whether the use of VM at the end of the activity sequence of the curriculum could have any impact on the knowledge gained through the use of PM in the previous parts of the curriculum. Could the presence of the VM, along with the unique affordance of the view of the charge flow, provide a conceptual framework (mental model) for VPP or PVP students that could scaffold knowledge acquisition in
the subsequent parts of the curriculum? Or could it provide a new conceptual framework for the PPV students that could prompt revision of the knowledge gained in the preceding parts of the curriculum? This study was situated within one instructional context; that is, the same instructors, method of instruction (inquiry), curriculum and teaching materials, and procedures (as defined by the Physics by Inquiry curriculum; e.g., students worked in small groups throughout the course) were used for all conditions.

Sample

The participants in the study were 194 undergraduates in a preservice elementary school teacher program (52 males, 142 females; \( M_{\text{age}} = 20.4 \) years, \( SD = 0.87 \)) enrolled in an introductory physics course at a university in Cyprus. The data were collected over a period of two semesters, because the number of students taking the course per semester was not enough to meet the needs of this study. First, we collected the data for the PPP, VVV, PVP, and PPV conditions and then the data for the VPP condition. For the first four conditions, all students from the course were randomly assigned to condition; these conditions had about the same number of participants: the PPP condition (38 students), the VVV condition (38 students), the PVP condition (39 students), and the PPV condition (39 students). For the VPP condition we randomly selected 40 students (\( M_{\text{age}} = 20.2 \) years, \( SD = 0.83 \)) from among the students taking the same course that the students in the other conditions had taken earlier. Additionally, within each condition students were randomly assigned to groups (of three) as suggested by the curriculum of the study.

We compared the different conditions on demographic characteristics. Conditions shared about the same proportion of men and women. Participants in the different conditions also (a) had about the same mean age, (b) had the same ethnicity (Greek Cypriots), (c) shared about the same K–12 educational background (e.g., all participants in all conditions were graduates of public schools, and all groups shared about the same distribution of participant scores on the university entrance exams), and (d) were at the same (second) year of their undergraduate studies when the data were collected. Finally, a one-way ANOVA showed that the participants’ pretest performance scores did not differ significantly between conditions, \( F < 1, \) ns (see Table 3 for means and standard deviations).

Curriculum Materials: Electric Circuits

For the purpose of this study, we used the first eight sections of the module of Electric Circuits of the Physics by Inquiry curriculum (McDermott & the Physics Education Group, 1996, p. 382–454), which we clustered in three parts as shown in Table 1. The rationale behind this clustering was based on the type and complexity of the electric circuits included in the sections, as well as the content that each part covered.

Part A (Sections 1 and 2) involves only basic circuits, namely one- and two-bulb circuits, and targets the development of a qualitative, conceptual model for electric circuits in the context of one- and two-bulb circuits. Specifically, in Section 1, the brightness of bulbs that are connected to a battery in different configurations is examined, and simple electric circuit concepts are introduced that will enable learners to account for relative brightness of the bulbs that they
TABLE 1
The Curriculum Covered in the Study

<table>
<thead>
<tr>
<th>Part A (Number of Experiments)</th>
<th>Part B (Number of Experiments)</th>
<th>Part C (Number of Experiments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Section 1: One-bulb circuits (9)</td>
<td>• Section 3: Extending the model for electric currenta (6)</td>
<td>• Section 6: Equivalent resistancec (3)</td>
</tr>
<tr>
<td>• Section 2: A model for electric currenta (3)</td>
<td>• Section 4: Series and parallel networksb (6)</td>
<td>• Section 7: Multiple batteriesc (11)</td>
</tr>
<tr>
<td>• Section 3: Extending the model for electric currentb (6)</td>
<td>• Section 5: Kirchoff’s first ruleb (3)</td>
<td>• Section 8: Kirchoff’s second ruleb, c (9)</td>
</tr>
</tbody>
</table>

Note. aThis section includes one- and two-bulb circuits. bThis section includes circuits with more than two bulbs. cThis section includes circuits with resistances and bulbs. Some of the experiments of Section 8 were further extended to include concepts studied in Part B (i.e., current).

observe. In Section 2, students are encouraged to construct a conceptual model for an electric circuit from direct experience with batteries and bulbs. In Part A, students are asked to connect a wire or a light bulb across a battery, which produces evidence that something is happening in the circuit. The wire becomes warm to the touch or the bulb glows. Given this evidence, along with the preexisting current-flow-based models that university students have constructed in their minds during the K–12 years (for examples of such models see Driver, Guesne, & Tiberghien, 1985; Driver, Squires, Rushworth, & Wood-Robinson, 1994), students are expected to use the concept of flow to explain and predict the behavior of electric circuits, without having to specify its nature (e.g., charge or electrons). According to the Physics by Inquiry curriculum developers, in Part A of the curriculum, it is enough for the students to know that the concept of flow is defined as something moving in the circuit and that it is called electric current. For the purposes of this study, students were officially introduced through the curriculum to the concept of flow and that it is called electric current at the beginning of Part B, to avoid any interference with the research design of the study. With this current-flow-based model, students are expected to envision the current flow as a continuous loop from one terminal of the battery, through the rest of the circuit, back to the other terminal of the battery, through the battery, and back around the circuit. Given that students have observed that a light bulb included in this circuit will light, they come to understand that the brightness of a bulb is an indicator of the amount of flow through the bulb. With this model, students are therefore expected to be able to predict which bulbs will light, which will be brightest, dimmest and so forth.

Part B (Sections 3 through 5) involves multibulb circuits with the bulbs connected in both series and parallel networks. These sections focus on (a) extending the qualitative model for electric current developed in Sections 1 and 2 to apply to more complex circuits, with series and parallel networks; and (b) transforming it into a semiquantitative current model by incorporating Kirchhoff’s first rule. More specifically, Sections 3 and 4 treat the behavior of bulbs in circuits with more than two bulbs and with different arrangements, respectively. In Section 5, a more quantitative analysis of electric circuits is introduced. The idea of current conservation is investigated by using ammeters to measure the magnitude of the current through the elements of a circuit.

Part C (Sections 6 through 8) involves circuits with multiple batteries and new resistive elements (nichrome wire resistances) in addition to bulbs. The focus of these sections is to further
extend the conceptual model developed in Parts A and B into a model for electric circuits that accounts for the concepts of resistance, voltage, and current. In particular, in Section 6, a quantified version of resistance is included, and learners are urged to investigate how the magnitude of the resistance of a network depends on the way in which its elements are connected to one another. Section 6 also encourages learners to look at how the total resistance of each network (in series or in parallel) relates to the current flow passing through it. In Section 7, a quantified version of voltage is introduced. In particular, the effect of adding batteries to a two-element (in series or in parallel) circuit is examined (batteries in a complete circuit produce voltages across the resistances in the circuit), and how this relates to the current passing through a circuit. Lastly, Section 8 incorporates more complicated circuits than the two-element circuits, and a more general quantitative conceptual model for electric circuits that accounts for multiple batteries, bulbs, and resistors is targeted. Overall, all eight sections of the Physics by Inquiry curriculum are organized in a dedicated path that is intended to support students in building gradually, through experimentation, a thorough conceptual model that predicts and explains the behavior of electric circuits.

Materials

Physical Material. PM experimentation involved the use of real instruments (voltmeter, ammeter) and objects (identical batteries, wires, switches and resistive elements, such as bulbs) in a conventional physics laboratory. During PM experimentation, feedback is available to the students through the behavior of the actual system (e.g., bulbs’ brightness) and through the instruments that are used to monitor the experimental setup (e.g., voltmeters, ammeters).

Virtual Material. VM experimentation involved the use of virtual instruments (voltmeter, ammeter) and objects (identical batteries, wires, switches and resistive elements, such as bulbs) to conduct the study’s experiments on a computer. In this study, Virtual Labs Electricity software (Riverdeep Interactive Learning, 2009) was used. Virtual Labs Electricity was selected because of its fidelity and the fact that it retained the same features and interactions of the domain of electric circuits as with PM. In this open-ended environment, students were able to design and test any Direct Current circuit mentioned in the curriculum by employing the “same” instruments and circuit parts (batteries, wires, switches and resistive elements, such as bulbs) as when experimenting with PM.

In the Virtual Labs Electricity environment, students were able to construct their own virtual experimental arrangements by simple and direct manipulation of virtual objects and virtual instruments. Circuits were created by clicking on icons representing electrical parts and moving the parts to the desired position in the circuit (see Figure 2). The software offered feedback during

---

1For a very few experiments from the curriculum, the software could not provide all the material needed for the experimental setup; hence, interactive simulations were developed and used to complement the Virtual Labs Electricity software. These simulations worked in the same manner as Virtual Labs Electricity (by clicking on icons representing electrical parts and moving the parts to the desired position in the circuit) and offered the same unique affordance (i.e., a view of the current flow).
FIGURE 2 The Virtual Labs Electricity screen.

the experiment by presenting information (e.g., the voltage across a resistive element, the amount of current) through the instrument displays of the software (when voltmeters and ammeters were included in a circuit) and by animating the phenomenon (e.g., bulbs’ brightness). This feedback was analogous to what is available to students through PM experimentation. In addition to this feedback, the VM provided feedback through instant notification when an electrical connection was made between two or more objects or circuit-parts (the electrical connection instantly changed color, which indicated that the connection was made). Finally, VM involved tools that could provide a schematic view of the circuit, as well as a view of the charge flow (dots moving at various speeds). The schematic view did not offer any additional information to the students using VM, because the schematic representation of the circuit was provided to all students in the Physics by Inquiry teaching material. In contrast, the view of charges was an affordance that was offered only by the VM.

Procedure

Within each condition, participants were assigned randomly to groups of three, which resulted in 13 groups for the PPP, VVV, PVP, and PPV conditions (for the PPP and VVV conditions we had 12 groups of three students and one of two students) and 14 groups for the VPP condition (12 groups
of three students and two of two students). In total, 66 groups were formed. All groups of all conditions worked in the same laboratory environment that hosted both conventional equipment and a computer network.

Before engaging with the learning material, students were introduced to the Physics by Inquiry curriculum and both PM and VM through a demonstration, regardless of their condition. The introduction to the Physics by Inquiry curriculum, routines and procedures was very important because they differ from those involved in the more traditional, passive modes of instruction that students experienced in physics courses during primary and secondary school. For example, the enactment of the Physics by Inquiry curriculum does not involve any lecturing, tutoring, or traditional textbook. On the contrary, the entire learning experience is based on collaborative inquiry, learning by doing, and experimentation. Students are seen as responsible for their own learning and are expected to construct knowledge collaboratively and to develop their understanding of physics concepts through conducting a carefully designed, structured sequence of inquiry-based experiments that are included in the Physics by Inquiry teaching materials.

Moreover, the role of the instructors in Physics by Inquiry is quite different from that of the instructor in traditional instruction. It is supportive in nature and requires instructors’ engagement in dialogues with the students in their groups at particular points (checkpoints) of the activity sequence, as specified by the Physics by Inquiry material. Through these dialogues, the instructors try to encourage reflection about the inquiry processes and practices involved in the activities of the Physics by Inquiry curriculum (e.g., constructing explanations, defending and challenging claims, interpreting evidence) and not to lecture or provide ready-made answers or solutions. There is no lecturing over the course of the learning process.

For consistency in the instructor–student interactions and dialogues, all instructors had meetings on a weekly basis that focused on reviewing the activities before these were encountered in class and considering possible issues to be discussed or prompts to be used during the dialogues with the students. For this study, the same seven instructors (one academic and six doctoral students) were always present for the PPP, VVV, PVP, and PPV conditions. Each instructor was assigned the same groups per condition throughout the study (two groups per condition were assigned to each of the doctoral student instructors, and one group per condition was assigned to the academic). For the VPP condition, we used the same seven instructors, with each instructor assigned to two groups. All instructors had been previously trained in implementing the Physics by Inquiry curriculum and had taught it for at least one academic year prior to this study.

A Complete Electric Circuit test was administered at the beginning and end of the study, with additional conceptual knowledge tests administered both before and after each part of the curriculum (A, B and C) to assess students’ understanding of concepts related to electric circuits as treated in the sections of that specific part (see Figure 1). The duration of the study was 15 weeks. Students met once a week for 1 hour and 30 minutes. The duration of the intervention, as well as the time-on-task, was the same for all conditions (28 hours and 30 minutes in total; see Table 2 for details).

DATA COLLECTION

Four different data sources were used, namely conceptual knowledge tests, instructors’ reflective journals, video data (including screen-captured data for all the conditions that included VM),
TABLE 2
The Study’s Time Framework

<table>
<thead>
<tr>
<th>Task</th>
<th>Time-on-Task for All Conditions Hours (Week of the Study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEC Test</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Test A</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Introduction of the Physics by Inquiry curriculum and demonstration of PM and VM</td>
<td>0.5 (2)</td>
</tr>
<tr>
<td>Intervention (Part A)</td>
<td>3 (3–4)</td>
</tr>
<tr>
<td>Test Ab</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Test Bb</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Intervention (Part B)</td>
<td>7.5 (5–9)</td>
</tr>
<tr>
<td>Test Bb</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Test Cb</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Intervention (Part C)</td>
<td>7.5 (10–14)</td>
</tr>
<tr>
<td>Test Cb</td>
<td>1 (14)</td>
</tr>
<tr>
<td>CEC Testb</td>
<td>2 (15)</td>
</tr>
<tr>
<td>Totalc</td>
<td>28.5 h</td>
</tr>
</tbody>
</table>

Note. aAll conditions were introduced to both VM and PM, whether their participants were going to use them or not during the study. They were also introduced to the Physics by Inquiry curriculum. bThe tests were completed at a prescheduled time outside the course. For this reason, the students who completed all the items on all of the study’s tests received a bonus toward their final grade in the course. cThe total time was the same for all conditions. Any student group in any condition that completed the experiments for a particular class meeting ahead of the others (usually students using VM) was asked to work on their laboratory notebook until the class meeting time was over. Students were asked to take brief notes during the class meeting, in order to avoid using up time on writing, and could elaborate on and complete their notes after finishing all of the experiments for that class meeting. They usually started on this at the end of a class meeting and finished at home.

and interviews. The conceptual knowledge tests were used to assess students’ understanding, and the instructors’ reflective journals and video data were used to gain insight into students’ experimentation processes. The interviews were used for triangulation and clarification purposes, and to gain deeper insight.

Conceptual Knowledge Tests

Paper-and-pencil tests were administered to assess students’ understanding of concepts concerning the electric circuits both before and after the study (Complete Electric Circuit test [CEC]). Additionally, (identical) tests specific to each part of the learning material were administered before and after each part (Tests A, B, and C; see Figure 1).

The items included on the conceptual knowledge tests were developed and used in previous research studies by the Physics Education Group of the University of Washington (e.g., McDermott & Shaffer, 1992; Shaffer & McDermott, 1992), as well as by our own research group (Zacharia, 2007). The tests for all parts (A, B, and C) were four open-ended items that asked conceptual
questions, all of which required explanations of reasoning. The CEC test included 13 open-ended items assessing all three parts (A, B, and C) of the study's curriculum (see the first column of the table in the Appendix for a sample item). This test targeted the specific concepts introduced in each part as well as the interconnections and interdependencies of these concepts. Items on the CEC test were different from items on Tests A, B, and C. All items on all tests consisted of several subitems (each subitem corresponded to one question). We always required an answer and an explanation or reasoning for each subitem.

Given that items on the CEC test and on the other tests were intended to be measuring the same concepts, internal consistency across parallel items was calculated by using Cronbach's alpha. For calculating Cronbach's alpha, the total score for each item on Tests A, B and C (an item on Test A, B, or C was comprised of several subitems) was compared to the total score of the corresponding sub-items on the CEC test. We used the data collected both at the pre- and posttest phase, separately. The subitems from Tests A, B, and C were not identical to the corresponding CEC test subitems but measured the same concepts. All alpha measures ranged from 0.71–0.86.

Instructors' Reflective Journals

All instructors were asked to keep a journal in which they had to document for each of their two assigned groups (a) any process-related difficulties or problems that the group faced while setting up and conducting an experiment (e.g., problems related to software, physical materials, curriculum, or collaboration), (b) any conceptual problem that the group encountered while conducting an experiment, and (c) students’ level of understanding of electric circuits concepts for each experiment from the curriculum material.

The documentation of these three aspects occurred in three phases. The first phase required the instructor to observe and document the process-related problems (e.g., problems related to building circuits through the use of PM or VM) that the students of each of his/her two groups faced while setting up and conducting an experiment over the course of a class meeting. Each instructor’s two groups were adjacent so that a single camcorder could videotape them and for the instructor to be able observe them at the same time. The discussions of each PM group were also audiotaped to provide additional documentation of the process-related problems encountered by the students. For the VM group, this was provided by the screen-captured data, which included audio data (for more details see the Video Data section below). The documentation of process-related problems in this phase was done mostly in real time. The videotaped and audiotaped material was used only when an instructor could not observe his/her groups directly, mainly due to commitments arising from the other two phases (described below).

The second phase involved documenting a process-related or a conceptual problem after an instructor was called unofficially (not for the purposes of a checkpoint) by a group of students to discuss a problem they faced while setting up or conducting an experiment, and they needed assistance with it in order to proceed with the experiment.

The third phase involved documenting a process-related or conceptual problem after an instructor was called officially (for the purposes of a checkpoint) by a group of students to discuss the experiments done since the previous checkpoint. As mentioned above, the Physics by Inquiry curriculum includes a number of prespecified checkpoints, which usually occur immediately after a new concept is introduced or at points where students had been found to need support.
In this phase, the instructors were required to use a rubric table that listed questions to be asked in order to determine each group’s level of understanding concerning the electric circuits concepts introduced in the experiments during the period covered by this checkpoint. To make the documentation easier, the electric circuits concepts introduced at each experiment were listed in the rubric table and were accompanied by a 3-point Likert scale indicating the level of the students’ understanding at a group level (not understood, partially understood, understood) for each electric circuit concept. The assigned instructor visited each group of students for each checkpoint throughout the study. When the instructors felt the need to supplement what they had documented in class during the second and third phase, they had the opportunity to use the videotaped and audiotaped material collected for each group after each class meeting.

For internal consistency reliability purposes, a second “silent” observer observed about 10% of the time spent by an instructor in either of his/her assigned groups per condition across the three phases. All instructors were compared against the data collected from a second silent observer. The data collected from the silent observer per instructor corresponded to approximately 10% of the data collected from each instructor in all conditions. These internal consistency data were collected both during class time and through the video and audio data collected. An experienced academic, who had prior multi-year experience with the Physics by Inquiry curriculum and its implementation, undertook the role of the silent observer. It should be noted that the silent observer did not participate in the instructional process and that the checkpoints he observed per instructor during class time were not necessarily the same checkpoints across all instructors. There are 36 checkpoints in the first eight sections of the electric circuits unit of the Physics by Inquiry curriculum. All internal consistency reliability measures (Cohen’s kappa) for all of the aspects documented in instructors’ reflective journals (i.e., process-related problems, understanding of electric circuits) were found to be at or above .84 across all instructors. In addition to the use of video data, the data from the instructors’ journals were also used as entry to a more detailed analysis of student actions during an experiment.

**Video Data**

Given the long duration of the study’s intervention (one semester), we only selected (randomly) two groups from each condition for analysis of their discourse and actions in order to identify whether students engage in different processes during PM or VM experimentation. The selection of these groups was done after students completed the CEC pretest. In particular, we randomly selected two groups per condition and compared the CEC pretest scores of the students of the selected groups from each condition to the scores of the remaining students in the same condition. This was to ensure that the students in the selected groups had similar levels of prior knowledge on electric circuits to the other student groups in that condition. We used the Mann-Whitney test and found no significant differences across all comparisons (\( p > .05 \)). This suggests that the groups selected were good representatives of their conditions in terms of relevant prior knowledge.

Although video and audio data were collected from each group throughout the study, for the selected groups, we also used a second camcorder at any period when they used PM. The goal was to have a better view of the lab bench at which the students were working with PM in order to capture better close-ups of student actions (e.g., building circuits) than with just the first camcorder. For VM, we used the screen-capture plus video–audio software (River Past Screen
Recorder Pro) to capture actual computer work activity (e.g., actions, sounds, movements that take place on the computer monitor).

After capturing student discourse and actions for each of the selected groups of each condition throughout the study, we intentionally selected and analyzed only certain episodes that involved the critical events that interested us (Powell, Francisco, & Maher, 2003). We used the instructors’ reflective journals to identify these critical events. From analysis of the instructors’ journals, we first identified the experiments for which the two selected groups for a condition appeared to deviate from the other groups of the other conditions, in terms of the problems they faced (process- or concept-related). We then located and isolated the (video) episodes that included the experiments that involved these critical events (points of differentiation across the groups) and proceeded with transcribing the corresponding dialogues and with coding students’ actions and activities (for PM data collected through camcorder and for VM data collected through the screen-capture software). The idea was to check whether these problems differentially affected students’ discourse and actions and therefore also affected the students’ processes in experimentation and their level of understanding of the electric circuits concepts introduced in these experiments. A total of about 1,090 minutes of student conversations were transcribed and coded. The corresponding actions of the students, within these 1,090 minutes of video, were also coded.

Interviews

For the purposes of this study, five students from each condition were randomly selected and interviewed, using a semi-structured interview protocol about two weeks after the teaching intervention ended. The protocol had two parts. However, for the purposes of this study, only the data from the first part were used. This part included questions that focused on students’ views, both positive and negative, of the sequence of experimental modes they experienced (PPP, VVV, VPP, PVP, PPV), the manipulatives they used (PM or VM), the instructional methods of the curriculum (e.g., working in groups, no lecturing), and the curriculum (per se) they used. For instance, they were asked if they liked the manipulative(s) they used as a means of experimentation and whether they identified any benefits or limitations related to the process of experimentation. The duration of this part of the interview was about 25 minutes.

DATA ANALYSIS

Conceptual Knowledge Tests

All tests were scored and coded blind to participant condition. We took the individual student as the unit of analysis. Each test consisted of items and each item of several subitems. The scoring of each subitem was performed through the use of a scoring rubric that included preset criteria (expected correct answer and expected correct explanation of reasoning; for an example, see the Appendix), which were used to score whether the elements of the participant’s overall response (answer and its accompanying reasoning) were correct. The scoring of the accompanying reasoning was based only on whether students provided specific concepts or evidence that were needed to support their answer, as prespecified in the scoring rubrics. A correct answer to a
subitem received one point, and its corresponding reasoning were scored in accordance with how many of its preset criteria were met. Each prespecified concept or evidence present in the reasoning received a half point. However, it should be noted that students received points only when they provided a correct answer and a corresponding correct or partially correct reasoning. No cases were found in which an incorrect answer and correct reasoning was provided. Students received no points for a correct answer accompanied by an incorrect reasoning. No matter how verbose some students were, they received no points unless they referred to the specific concepts or evidence that needed to be included in their explanations of reasoning.

The maximum score for each subitem varied according to the number of prespecified elements required to be present. However, the maximum score for an item (the sum of the maximum scores of its subitems) did not vary significantly across the items on a test. Hence, the scales for the items on a test were about the same. An individual's total score on a test was derived by adding all the assigned subitem scores, those for answers and for explanations and reasoning, and by adjusting it to fit on a 100-point scale. The adjustment to a 100-point scale aimed at making it easier to compare scores between different tests. Two independent coders reviewed about 20% of the data. The reliability measures (Cohen’s kappa) for scoring of the CEC test and Tests A, B, and C were .89, .92, .91, and .88, respectively.

The quantitative analysis involved one-way ANOVA for comparing the pretest scores for the five conditions on each test (CEC test and Tests A, B, and C), paired samples t-test for comparing the pretest–posttest scores for each condition across all tests, and one-way ANCOVA for comparing the posttest scores for the five conditions on each test (CEC test and Tests A, B, and C). For the last procedure, the students' scores in the corresponding pretests were used as the covariate. The aim of the first procedure was to determine whether the five conditions of the study were comparable with regard to the sample’s entry understanding of relevant physics concepts from the subject domain of electric circuits, before the study and before each part. The aim of the second procedure was to investigate whether each condition improved students' conceptual understanding, within the context of the Physics by Inquiry curriculum. The aim of the third procedure was to investigate whether the five conditions of the study resulted in different outcomes in terms of understanding of physics concepts in the domain of electric circuits, as measured by each test.

Along with the quantitative analysis of the conceptual knowledge tests, we also did a more qualitative analysis that focused on identifying and classifying students’ scientifically acceptable (SAC) or not scientifically acceptable (NSAC) conceptions concerning electric circuits within their answers on the tests. This analysis followed the procedures of open coding (Strauss & Corbin, 1998), in which the researchers first underlined the most important sentences in each student’s pre- and posttests and marked keywords that characterized the student’s conceptions with respect to behavior of electric circuits, measurement of current, resistance, and voltage. By comparing the sentences underlined and the keywords derived from the tests, the content-specific similarities and differences in students’ test responses about the aforementioned four categories of conceptions (behavior of electric circuits, measurement of current, resistance, and voltage) were explored and summarized.

Then, the researchers constructed qualitatively different subcategories of description, across rather than within the responses, that were used to classify the conceptions of behavior of electric circuits, measurement of current, resistance, and voltage held by students for each condition separately. By comparing the similarities and differences between the students of each condition,
subcategories of conceptions emerged for each of the aforementioned four categories of conceptions (for examples of such subcategories, see Table 7). Each subcategory was intended to show a unique way of understanding the phenomenon being studied. Therefore, the purpose of the open coding analysis was to reveal the subcategories of description that could characterize the qualitatively different perspectives in which behavior of electric circuits, measurement of current, resistance, and voltage were conceptualized or experienced by the students of each group. As a result of this analysis, we created a table for each subcategory, giving one SAC and its corresponding NSACs (for an example, see Table 8). After completing this process, we systematically compared all of these tables to identify whether certain patterns exist within a condition on both the pretest and the posttest (e.g., whether a dominant NSAC exists on a pretest and whether this NSAC remains dominant on the posttest), as well as between conditions on both the pretest and the posttest (e.g., whether the students of different conditions share the same NSAC on either the pretest or the posttest).

In addition, the prevalence for each of the resulting subcategories for each test was calculated, as well the mean frequencies and standard deviations of SAC and NSAC in the five conditions and on the four tests at pre- and posttest (see Table 9). The aim of the latter calculation was to compare whether students’ conceptions changed over the course of the study. This procedure was essential because it clarified whether students with similar scores also shared the same ideas, either SAC or NSAC conceptions.

For internal consistency reliability purposes, a second independent rater reviewed about 20% of the data. The reliability measures (Cohen’s kappa) for identifying subcategories of SAC and NSAC as described above for the CEC test and Tests A, B, and C were .82, .87, .84, and .80, respectively. The reliability measures (Cohen’s kappa) for classifying students’ conceptions according to the resulting subcategories for the CEC test and Tests A, B, and C, for another 20% of the initial data, were .93, .94, .90, and .90, respectively.

Finally, we proceeded with a second qualitative analysis, in which we identified the conceptual models that the students developed for explaining and predicting the behavior of electric circuits before the study and after each part of the curriculum. For this purpose, we used the CEC pretest and the posttests A, B, and C and followed the procedures of open coding (Strauss & Corbin, 1998) for analyzing the data. As mentioned in the Curriculum Materials section, the unit on electric circuits in the Physics by Inquiry curriculum supports the development of a current-flow-based conceptual model. Therefore, to identify the conceptual models that the students in each condition developed and used for explaining and predicting the behavior of electric circuits, first we underlined the most important sentences and marked some keywords that characterized the student’s conceptual model with respect to the flow of current in an electric circuit for each part of curriculum separately. By comparing the sentences underlined and the keywords derived from the tests for each part, the content-specific similarities and differences between students’ test responses about the current flow in an electric circuit were explored and summarized. Second, we constructed qualitatively different categories of description to classify the conceptual current flow model held by students at each part of the curriculum for each condition separately. Each category was intended to show a unique way of describing the current flow in a circuit (for the most prevalent categories identified, see Figure 3).

We further calculated the prevalence for each conceptual model and created a flowchart showing the conceptual model paths that the students in each condition followed from the beginning to the end of the study (see Figure 4). This type of analysis was performed in order to examine
whether the use of PM or VM, in the three different parts of the curriculum, enabled students to
develop different conceptual models and therefore to explain and predict the behavior of elec-
tric circuits in a different manner. It should be noted that these current-flow-based conceptual
models do not reflect students’ conceptual understanding holistically. This is because they do not
account for all of the concepts associated with electric circuits, such as the rules depicting how
bulbs or resistances and/or networks of bulbs or resistances are connected, Kirchhoff’s second
rule (which is built on a voltage-based conceptual model), and so forth. We selected this type
of analysis (current-flow-based conceptual models) because it was the one under emphasis for
promoting students’ conceptual understanding in our study’s curriculum (see the Curriculum
Materials section).

For internal consistency reliability purposes, a second independent rater reviewed about 20%
of the data. The reliability measures (Cohen’s kappa) for identifying and classifying students’
mental models concerning electric circuits for the CEC pretest and posttests A, B, and C were
.91, .88, .86, and .87, respectively. The reliability measures for calculating the prevalence of
each conceptual model for the CEC pretest and posttests A, B, and C were .89, .93, .90 and .88,
respectively.

Instructors’ Reflective Journals

For the journal data analysis, we again followed open coding from grounded research methodology
(Strauss & Corbin, 1998) but took the group as the unit of analysis. In particular, we developed
codes accounting for the process-related problems that the students of each group of each condition
encountered during experimentation. Data for all groups from each condition were analyzed. Two
coders independently completed all coding; Cohen’s kappa for the initial coding was .92, and
differences in the assigned codes were resolved through discussion. After finalizing the coding
scheme, we grouped different codes together based on what a problem involved, resulting in
two different categories. The first category included problems (process-related in nature) that
cconcerned the feedback received by the manipulatives used, particularly by PM (e.g., PM did not
provide observable feedback in some circuits). The second category included problems, again
process-related, that concerned the construction of complex circuits (e.g., students using PM
faced problems with constructing circuits that included networks of bulbs—branches of bulbs
connected in parallel—that were connected in series).

We did not break the conceptual problems down into more specific categories, as reported
by the instructors in their journals, but only coded whether a group of students faced con-
ceptual problem(s) in general or not for each experiment. We decided to use the conceptual
knowledge test data for a detailed analysis of students’ conceptual problems. We judged that
the conceptual knowledge tests were rich enough to unveil students’ SAC or NSAC con-
ceptions. Given the two categories of process-related problems described and the broad cate-
gory of conceptual-related problems, the initial list of codes was recoded to account for these
three categories, as well as for cases for which no problems (process- or conceptual-related)
were encountered. For the process-related problems, we further separated and coded for stu-
dents who requested support from the instructor and those who managed to handle these
problems themselves. Each experiment was assigned a new code that corresponded to one
or more of these categories. Two coders independently completed all coding; Cohen’s kappa
These data were then quantitatively treated in two ways. First, we used a chi-square goodness-of-fit test to examine whether the groups in each condition faced the same problems (or no problems) for each of the study’s experiments. Second, we used a chi-square test (Cramer’s V) to examine whether the study’s five conditions differed in terms of the problems that their groups faced for each of the study’s experiments.

Finally, we coded for the instructors’ assessment of the level of students’ understanding for each electric circuit concept involved in an experiment. In this case, we assigned a code to each concept (0 = not understood, 1 = partially understood, 2 = understood), and then we calculated a mean score for each experiment for each student group. This score was then used to check on the possibility that the problems faced by the students while conducting the experiments for each part of the curriculum were associated with their levels of understanding of the concepts involved in those experiments. To check for a possible association of condition, problems faced by the students and levels of understanding of the students, we used a two-way ANOVA.

Video Data

For the video, audio, and screen-captured data analysis, we also followed open coding from grounded research methodology (Strauss & Corbin, 1998) and took the group as the unit of analysis. In the case of the audio material, we transcribed the selected conversations and coded what the students were talking about (e.g., about the experimental setup, about their observations). After all transcribed conversations were coded and the list of codes was finalized, all coded transcripts were reviewed once more for consistency reasons. Interreliability data were collected as well. A second coder who did not have access to the first round of coding repeated the whole coding process. Cohen’s kappa was calculated at .88. Differences in the assigned codes were resolved through discussion.

For analyzing the data related to what the students were doing with PM or VM, we coded the video and screen-captured data for students’ actions (e.g., building a circuit, playing with the material, repeating an experiment). The codes emerged through open coding and aimed at capturing the students’ actions during their work with the PM or VM. After the list of codes was finalized, all coded transcripts were also reviewed once more for consistency purposes. Interreliability data were collected as well. A second coder who did not have access to the first round of coding repeated the whole coding process. Cohen’s kappa was calculated at .84. Differences in the assigned codes were resolved through discussion.

The resulting codes for both discourse and actions were combined and presented in timeline graphs, following the approach of Schoenfeld (1989). The x-axis of the graph displays time, and the y-axis displays what the students were talking about and their actions related to PM or VM usage (see Figure 5). The goal was to reveal through these graphs the interrelationships of the codings (student talk and actions) over time. Overall, we made such graphs for each experiment for which the selected groups of a particular condition appeared to deviate from the selected groups of the other conditions in terms of the problems they faced. We produced a graph for each selected group from each condition for each experiment. We then compared the resulting graphs for each experiment across the five conditions to identify similarities.
PHYSICAL AND VIRTUAL MANIPULATIVES

and differences in the combinations of codes over time and proceeded with the corresponding interpretations.

Interviews

The interview analysis focused solely on triangulating the findings from the analysis of the instructors’ journals. We therefore focused only on the problems that the students mentioned encountering during the study that were related to the sequence of experimental modes they experienced (PPP, VVV, VPP, PVP, PPV), the manipulatives they used (PM or VM), the instructional methods of the curriculum (e.g., working in groups, having multiple instructors, no lecturing), and the curriculum material (per se) that they used.

The codes for the transcribed interviews were developed by following open coding procedures from grounded research methodology (Strauss & Corbin, 1998) and accounted for the problems that the students (we took the individual student as the unit of analysis) reported they encountered across the indicated categories. Specifically, we first isolated the segments at which the students made explicit reference to a problem that concerned any of these four categories and transferred them into an Excel file by placing each utterance in a different cell. Each student utterance was analyzed separately and received only one code. After the list of codes was finalized, all coded transcripts were reviewed once more, to check for consistency in the applied coding. Another coder who did not have access to the first analysis then repeated the coding (Cohen’s kappa = .94). Differences in the assigned codes were resolved through discussion. From analysis, codes emerged only for problems students encountered during the use of PM, because no problems were reported concerning VM use or the other possible categories. The final codes identified from the transcribed interviews perfectly matched the codes from the analysis of the instructors’ journals. Both codes concerned process-related problems; the first code accounted for problems related to problematic feedback, and second code accounted for problems related to the construction of complex circuits.

RESULTS

Conceptual Knowledge Acquisition

The one-way ANOVA procedure indicated that the conditions did not differ in pretest scores on any tests, $F < 1$, ns. Mean scores and standard deviations are shown in Table 3.

The paired samples $t$-test showed that all five conditions improved students’ conceptual understanding for each part of the curriculum and for the curriculum as a whole ($p < .001$ for all comparisons, which is lower than the .0025 (0.05/20), the lowest $p$-value given by the Holm–Bonferroni method; see Table 4).

An ANCOVA was done for students’ scores on each of the posttests, with pretest score as covariate and condition as between-subjects factor. For Test A, the ANCOVA (with Pretest A as a covariate) did not reveal a main effect of condition, $F(4, 188) = 0.5, p = .73$, partial $\eta^2 = .01$ (see Table 5). This finding suggests that the use of either PM or VM was equally effective in promoting students’ understanding of the electric circuits concepts introduced in Part A across all conditions.
TABLE 3
Mean Scores and Standard Deviations on Each of the Tests for All Five Conditions

<table>
<thead>
<tr>
<th>Group</th>
<th>Test A</th>
<th>Test B</th>
<th>Test C</th>
<th>CEC Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>PPP pretest</td>
<td>30.7  (8.2)</td>
<td>15.2  (9.0)</td>
<td>20.4  (7.3)</td>
<td>18.0  (8.1)</td>
</tr>
<tr>
<td>PPP posttest</td>
<td>68.5  (9.0)</td>
<td>58.8  (12.7)</td>
<td>66.0  (10.7)</td>
<td>61.6  (11.1)</td>
</tr>
<tr>
<td>VVV pretest</td>
<td>29.3  (7.4)</td>
<td>15.5  (9.1)</td>
<td>19.7  (7.5)</td>
<td>17.4  (8.4)</td>
</tr>
<tr>
<td>VVV posttest</td>
<td>69.0  (9.2)</td>
<td>69.1  (11.0)</td>
<td>77.1  (11.2)</td>
<td>76.5  (11.9)</td>
</tr>
<tr>
<td>VPP pretest</td>
<td>31.4  (7.9)</td>
<td>16.2  (9.7)</td>
<td>21.1  (7.2)</td>
<td>19.1  (8.3)</td>
</tr>
<tr>
<td>VPP posttest</td>
<td>70.5  (8.6)</td>
<td>60.0  (13.3)</td>
<td>67.1  (11.3)</td>
<td>62.3  (11.4)</td>
</tr>
<tr>
<td>PVP pretest</td>
<td>32.0  (6.8)</td>
<td>16.0  (10.0)</td>
<td>21.8  (8.3)</td>
<td>19.0  (9.4)</td>
</tr>
<tr>
<td>PVP posttest</td>
<td>71.3  (10.2)</td>
<td>68.4  (12.0)</td>
<td>75.9  (12.2)</td>
<td>74.3  (12.4)</td>
</tr>
<tr>
<td>PPV pretest</td>
<td>31.1  (6.9)</td>
<td>16.8  (8.3)</td>
<td>21.1  (6.4)</td>
<td>18.6  (7.7)</td>
</tr>
<tr>
<td>PPV posttest</td>
<td>69.5  (9.5)</td>
<td>60.3  (9.7)</td>
<td>71.6  (9.0)</td>
<td>66.8  (9.1)</td>
</tr>
</tbody>
</table>

Note. aThe pretest and posttest versions of each test were identical. bScores were adjusted to a 100-point scale. This means that the scores reported here can be interpreted as percentage of the maximum score.

For Test B, the ANCOVA revealed a main effect of condition, $F(4, 188) = 28.7$, $p < .001$, partial $\eta^2 = .38$, and of Pretest B scores (covariate) on students’ Posttest B scores, $F(1, 188) = 563.3$, $p < .001$, partial $\eta^2 = .75$, but no interaction between condition and pretest scores, $F < 1$, ns. Bonferroni-adjusted pairwise comparisons revealed that students’ Posttest B scores in the VVV and PVP conditions were significantly higher than those of the students in the PPV, VPP,

TABLE 4
The Paired Samples T-Test Results for Each of the Tests

<table>
<thead>
<tr>
<th>Conceptual Tests</th>
<th>Condition</th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest A–Pretest A</td>
<td>PPP</td>
<td>34.71</td>
<td>37</td>
<td>&lt;.0001</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>VVV</td>
<td>44.56</td>
<td>37</td>
<td>&lt;.0001</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>VPP</td>
<td>37.13</td>
<td>39</td>
<td>&lt;.0001</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>PVP</td>
<td>40.60</td>
<td>38</td>
<td>&lt;.0001</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>39.34</td>
<td>38</td>
<td>&lt;.0001</td>
<td>4.6</td>
</tr>
<tr>
<td>Posttest B–Pretest B</td>
<td>PPP</td>
<td>38.20</td>
<td>37</td>
<td>&lt;.0001</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>VVV</td>
<td>56.64</td>
<td>37</td>
<td>&lt;.0001</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>VPP</td>
<td>40.45</td>
<td>39</td>
<td>&lt;.0001</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>PVP</td>
<td>62.19</td>
<td>38</td>
<td>&lt;.0001</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>52.47</td>
<td>38</td>
<td>&lt;.0001</td>
<td>4.8</td>
</tr>
<tr>
<td>Posttest C–Pretest C</td>
<td>PPP</td>
<td>34.67</td>
<td>37</td>
<td>&lt;.0001</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>VVV</td>
<td>50.60</td>
<td>37</td>
<td>&lt;.0001</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>VPP</td>
<td>36.43</td>
<td>39</td>
<td>&lt;.0001</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>PVP</td>
<td>46.08</td>
<td>38</td>
<td>&lt;.0001</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>52.99</td>
<td>38</td>
<td>&lt;.0001</td>
<td>6.4</td>
</tr>
<tr>
<td>CEC Posttest–CEC Pretest</td>
<td>PPP</td>
<td>35.17</td>
<td>37</td>
<td>&lt;.0001</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>VVV</td>
<td>51.47</td>
<td>37</td>
<td>&lt;.0001</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>VPP</td>
<td>38.2</td>
<td>39</td>
<td>&lt;.0001</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>PVP</td>
<td>46.96</td>
<td>38</td>
<td>&lt;.0001</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>56.01</td>
<td>38</td>
<td>&lt;.0001</td>
<td>5.7</td>
</tr>
</tbody>
</table>
TABLE 5
One-Way ANCOVA Results and Effect Sizes of the Post Hoc Comparisons Across Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Pairwise Comparisons</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A</td>
<td>0.5</td>
<td>4, 188</td>
<td>= .73</td>
<td>PPP - VVV</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - VPP</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PVP</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - VPP</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PVP</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PPV</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PVP</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PPV</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PPV</td>
<td>0.17</td>
</tr>
<tr>
<td>Test B</td>
<td>28.7</td>
<td>4, 188</td>
<td>&lt; .001</td>
<td>PPP - VVV*</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - VPP*</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PVP*</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PPV*</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - VPP*</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PVP*</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PPV*</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PVP*</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PPV*</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPV - PPV*</td>
<td>1.49</td>
</tr>
<tr>
<td>Test C</td>
<td>19.5</td>
<td>4, 188</td>
<td>&lt; .001</td>
<td>PPP - VVV*</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - VPP*</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PVP*</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PPV*</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - VPP*</td>
<td>1.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PVP*</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PPV*</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PVP*</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PPV*</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPV - PPV*</td>
<td>0.48</td>
</tr>
<tr>
<td>CEC Test</td>
<td>40.37</td>
<td>4, 188</td>
<td>&lt; .001</td>
<td>PPP - VVV*</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - VPP*</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PVP*</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP - PPV*</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - VPP*</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PVP</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VVV - PPV*</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PVP*</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VPP - PPV*</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPV - PPV*</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Note. *The mean difference is significant at the .01 level; we have used underlining to signal the condition that did better in each comparison where there was a significant difference between the conditions.

and PPP conditions (p < .01 for all statistically significant differences). The pairwise comparisons did not show a significant difference between the students’ Posttest B scores in the PVP and VVV conditions and between the PPP, VPP, and PPV conditions.

For Test C, the ANCOVA also revealed a main effect of condition, $F(4, 188) = 19.5$, $p < .001$, partial $\eta^2 = .29$, and of Pretest C scores (covariate) on students’ Posttest C
TABLE 6
Comparison of Test Performance Across All Conditions

<table>
<thead>
<tr>
<th>Test</th>
<th>Comparisons Among Conditions</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A</td>
<td>$\text{VVV} = \text{PVP} = \text{PPV} = \text{PPP}$</td>
<td>No real advantage for PM or VM in Part A.</td>
</tr>
<tr>
<td>Test B</td>
<td>$(\text{VVV} = \text{PVP}) &gt; (\text{PPV} = \text{PPP})$</td>
<td>VM outperforms PM in Part B.</td>
</tr>
<tr>
<td>Test C</td>
<td>$(\text{VVV} = \text{PVP}), (\text{VVV} &gt; \text{PPV}), \quad (\text{PPV} = \text{PVP}) &gt; (\text{VPP} = \text{PPP})$</td>
<td>PM can be as effective as VM in Part C, but only when preceded by VM in Part B.</td>
</tr>
<tr>
<td>CEC test</td>
<td>$(\text{VVV} = \text{PVP}) &gt; \text{PPV} &gt; (\text{VPP} = \text{PPP})$</td>
<td>The presence of VM in a PM-dominant curriculum is important for Parts B and C, with Part B being the most crucial.</td>
</tr>
</tbody>
</table>

Note. *Based only on posttest results.

scores, $F(1, 188) = 258.2, p < .001$, partial $\eta^2 = .58$, but no interaction between condition and pretest scores, $F < 1$, ns. Bonferroni-adjusted pairwise comparisons revealed that students’ Posttest C scores in the VVV, PVP, and PPV conditions were significantly higher than scores in the VPP and PPP conditions ($p < .01$ for all statistically significant differences). The pairwise comparisons also showed that the VVV condition had significantly higher Posttest C scores than did PPV students, but did not differ from those in the PVP condition. Additionally, the pairwise comparisons did not show a significant difference between the Posttest C scores of students in the PVP and PPV conditions or between the PPP and VPP conditions.

Finally, for the CEC test, the ANCOVA also revealed a main effect of condition, $F(4, 188) = 40.37, p < .001$, partial $\eta^2 = .46$, and of CEC pretest scores, $F(1, 188) = 342.01, p < .001$, partial $\eta^2 = .64$, but no interaction between condition and CEC pretest scores, $F < 1$, ns. Bonferroni-adjusted pairwise comparisons revealed that students’ CEC posttest scores were higher in the VVV and PVP conditions than in the PPV, VPP, and PPP conditions ($p < .01$ for all statistically significant differences; the effect sizes for these post hoc comparisons are shown in Table 5). Within these three conditions, students in the PPV condition scored higher than the students in the VPP and PPP conditions, who had the lowest scores overall. Pairwise comparisons did not show a significant difference between the students’ CEC posttest scores in the VVV and PVP conditions and between the students’ CEC posttest scores in the VPP and the PPP conditions. A summary of the results is presented in Table 6.

The overall picture suggested by these comparisons is that VM emerges out as more effective than PM. With regard to the different parts of the curriculum, we see that in Part A of the curriculum, there is no real advantage for PM or VM. In Part B, the students using VM outperform PM students, while in Part C, PM can be as effective as VM but only when preceded by VM in Part B.

Understanding of Electric Circuits Concepts

As described under data analysis, we further analyzed students’ answers on the different tests to reveal their qualitative understandings. We identified for each test conceptions of behavior
of electric circuits, measurement of current, resistance, and voltage. These four categories of conceptions consisted of a number of subcategories (for examples of such subcategories, see Table 7), each of which included one SAC and a number of NSACs that students could hold.

In Table 8, we present an example of our analysis. Table 8 shows the SAC for Kirchhoff’s second rule (subcategory of the measurement of voltage category) and all the different NSACs we identified. The table displays the percentage of students holding the different conceptions on the Pretest C and Posttest C for all conditions. The same process was followed for each SAC for the pre- and posttest for each part (A, B and C). Specifically, for each SAC we identified the related NSACs and calculated the corresponding prevalence for each condition separately. We then compared the SAC and NSACs for each pretest and posttest across conditions and compared the pretest to the corresponding posttest for each condition separately to investigate whether there were similar or different patterns within and between conditions.

For Test A, we found the students in all conditions to share the same SACs and most prevalent NSACs on both pretest and posttest (for examples of SACs and prevalent NSACs, see Table 7). In fact, the prevalence of each SAC and its corresponding most prevalent NSACs on Pretest A and Posttest A were found to be about the same for all conditions. We also found that the change in prevalence of these SACs and NSACs from Pretest A to Posttest A followed the same pattern. In particular, SAC prevalence was found to increase while the presence of the most prevalent NSACs decreased in the same manner in all conditions. It should be noted that both modes of experimentation (PM or VM) were equally effective in enabling students to overcome their NSACs and adopt the SACs in Part A, which implies that using PM or VM (along with its additional affordance, i.e., view of the charge flow) in a context similar to that of Part A did not differentially affect students’ conceptual understanding.

For Test B, we found that on the pretest students in all conditions shared the same SACs and most prevalent NSACs and showed similar prevalence for each SAC and corresponding NSACs. However, on Posttest B, we found that for students in the conditions using VM in Part B (PVP and VVV), each SAC was more prevalent and each corresponding NSAC less prevalent than in the conditions that used PM (PPP, VPP, PPV). While we found each SAC and its corresponding NSACs to change prevalence from Pretest B to Posttest B in every condition, the change in prevalence was more dramatic for the conditions that used VM in Part B than the conditions that used PM. PVP and VVV conditions were found to increase in prevalence of SACs and to decrease in prevalence of NSACs more than did the PPP, VPP, and PPV conditions. Compared to the findings for Part A, the use of VM, along with its unique affordance (view of the charge flow), appears to be more effective in enabling students to overcome their NSACs and adopt the SACs in Part B, which indicates that the use of PM and VM differentially affected students’ conceptual understanding in the context of Part B.

For Test C, as for Test B, we found that on the pretest, all conditions shared the same SACs and most prevalent NSACs, as well as having similar prevalence for each SAC and corresponding NSACs. For Posttest C though, the results differed from those found in Posttest B. This time, the PVP condition, whose students used PM in Part C, was found to follow a similar pattern as the PPV and VVV conditions, whose students used VM in Part C. In other words, the PPV, VVV, and PVP conditions were found to share the same SACs and most prevalent NSACs and have similar prevalence of each SAC and corresponding NSACs, on both the pretest and the posttest for Part C. Moreover, for the PPV, VVV, and PVP conditions, there was greater prevalence of each SAC and lower prevalence of each of its corresponding NSACs than for the
### TABLE 7
Examples of Subcategories of Student Conceptions of Behavior of Electric Circuits, Measurement of Current, Resistance, and Voltage and the Corresponding SAC and Most Prevalent NSAC

<table>
<thead>
<tr>
<th>Categories of Conceptions</th>
<th>Examples of Sub-Categories of Conceptions</th>
<th>SAC</th>
<th>Most Prevalent NSAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior of electric circuits</td>
<td>Complete/closed single-bulb electric circuit</td>
<td>A complete single-bulb electric circuit is a circular, closed route arrangement of a bulb, a battery, and a wire, in which each of the two terminals of the bulb is connected with a different terminal of the battery (see the figures below). In this case, current is passing through all circuit elements and the bulb lights.</td>
<td>A complete single-bulb electric circuit is a circular, closed route arrangement of a bulb, a battery, and a wire, in which one of the terminals of the bulb is connected with both of the battery’s terminals (see the figures below). In this case, current is passing through all circuit elements and the bulb lights.¹</td>
</tr>
</tbody>
</table>

| Measurement of current | Kirchhoff’s first rule | The total current out of a node is equal to the total current into the node, or The algebraic sum of the currents at a node is zero. | The total current out of a node is not the same as the total current into the node.² |

| Measurement of resistance | Equivalent resistance: Addition of resistances in parallel | The addition of resistances in parallel decreases the total resistance of the resulting parallel network of resistances. | The addition of resistances in parallel increases the total resistance of the resulting parallel network of resistances.³ |

| Measurement of voltage | Multiple batteries: Addition of batteries in series | The addition of batteries in series increases the voltage and the current through the circuit only when the total resistance of the circuit remains the same, or As the resistance in a circuit increases, the voltage must be increased for the current through the circuit to remain the same. | The addition of batteries in series increases the voltage and the current through the circuit (no reference is made to the total resistance of the circuit).⁴ |

---

*Note.* ¹This NSAC was identified on the CEC pretest and pretest A. It was not found on the CEC posttest and posttest A. ²This NSAC was most prevalent on the CEC and Test A pre- and posttests. Its prevalence decreased, in some conditions more than others, from pre- to posttest. ³This NSAC was most prevalent on the CEC and Test B pre- and posttests. Its prevalence decreased, in some conditions more than others, from pre- to posttest. ⁴This NSAC was most prevalent on the CEC and Test C pre- and posttests. Its prevalence decreased, in some conditions more than others, from pre- to posttest.
TABLE 8
Prevalence of Students’ SACs and NSACs About Kirchhoff’s Second Rule (From the Qualitative Analysis)

<table>
<thead>
<tr>
<th>Conceptions</th>
<th>PPP (n = 38)</th>
<th>VVV (n = 38)</th>
<th>VPP (n = 40)</th>
<th>PVP (n = 39)</th>
<th>PPV (n = 39)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
</tr>
<tr>
<td>SAC^a</td>
<td>0 (0)</td>
<td>44.7 (17)</td>
<td>0 (0)</td>
<td>81.5 (31)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>NSAC^b 1^c</td>
<td>89.4 (34)</td>
<td>55.2 (21)</td>
<td>92.1 (35)</td>
<td>18.4 (7)</td>
<td>90.0 (36)</td>
</tr>
<tr>
<td>NSAC 2^d</td>
<td>42.1 (16)</td>
<td>0 (0)</td>
<td>28.9 (11)</td>
<td>0 (0)</td>
<td>30.0 (12)</td>
</tr>
<tr>
<td>NSAC 3^e</td>
<td>36.8 (14)</td>
<td>0 (0)</td>
<td>50.0 (19)</td>
<td>0 (0)</td>
<td>42.5 (17)</td>
</tr>
<tr>
<td>NSAC 4^f</td>
<td>65.7 (25)</td>
<td>39.4 (15)</td>
<td>76.3 (29)</td>
<td>13.1 (5)</td>
<td>75.0 (30)</td>
</tr>
<tr>
<td>NSAC 5^g</td>
<td>31.5 (12)</td>
<td>15.7 (6)</td>
<td>26.3 (10)</td>
<td>0 (0)</td>
<td>25.0 (10)</td>
</tr>
<tr>
<td>NSAC 6^h</td>
<td>44.7 (17)</td>
<td>31.5 (12)</td>
<td>39.4 (15)</td>
<td>0 (0)</td>
<td>45.0 (18)</td>
</tr>
<tr>
<td>NSAC 7^i</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>2.5 (1)</td>
</tr>
</tbody>
</table>

Note. ^aSAC denotes Scientifically Acceptable Conception. The sum of the voltages across all the elements along any path or branch is the same for all paths between two nodes, or the voltage across the battery in a current loop is equal to the sum of the voltages across the other elements. The students who adopted SACs were counted only in the SAC category. ^bNSAC denotes Not Scientifically Acceptable Conception. Students who did not adopt the SACs stated all NSACs. In addition, the response of one individual student may appear in more than one NSAC category depending on the NSACs that were evident in the response. ^cNSAC 1: The sum of the voltages across all the elements along any branch is not the same for all branches between two nodes. ^dNSAC 2: The voltage of each circuit element in a current loop (across the battery) is the same. ^eNSAC 3: The voltage of a branch between two nodes is the same as the sum of the voltages of all the elements along that branch. ^fNSAC 4: The voltage of branches between two nodes is the same as the voltage of parallel branches that are not connected directly across the two nodes. ^gNSAC 5: When two or more parallel networks of elements are connected in series, the voltage across any of these networks is always the same. ^hNSAC 6: The voltage across the elements or the networks of elements along any branch is the same as the voltage across identical elements or networks of elements for all branches between two nodes. ^iNSAC 7: The voltage across the battery of a circuit is equal to the sum of the voltages across all the elements of the circuit.  The percentage and number (included in the parentheses) refer to students who explicitly mentioned the particular conception referred to, which does not mean that other students might not also share these conceptions. In other words, it is very possible that certain conceptions are more widespread than the numbers suggest.
PPP and VPP conditions on posttest C (for an example see Table 8). This means that the students’ conceptual understanding was differentially affected, but we cannot conclude that it was because of VM and its advantageous affordance because PVP had the same effect on students’ learning as the VVV and PPV conditions. What is interesting about this latter finding is that the presence of the advantageous VM does not appear to be necessary for learning from experimenting in Part C.

The qualitative analysis of the CEC test revealed exactly the same SACs and NSACs as on Tests A, B, and C, which followed the same patterns both within and between groups as those described above. The CEC test findings served to triangulate the findings for Tests A, B, and C, which add further credence to our qualitative analysis of students’ conceptions.

Table 9 shows the overall picture by means of mean frequencies of SAC and NSAC on the four tests at the pre- and posttest in the five conditions. Overall, the results of the qualitative analysis confirm what was found in the quantitative analysis. On the CEC pretest and the pretests of the different parts (A–C), very few students already possessed correct conceptions of the domain, and they displayed a variety of conceptions that were not scientifically acceptable. On each of the posttests, a higher number of correct conceptions could be seen than on the corresponding pretests, and in all cases, the number of NSACs found in the answers of the students went down. For Test A, no differences between conditions could be found with regard to change of NSAC to SAC. The type of manipulation (VM or PM) therefore did not influence students’ qualitative changes on Test A. For Test B, students in the VVV and PVP conditions shifted from NSACs to SACS to a greater extent than did those in the PPV, VPP, and PPP conditions, which implies that for Test B, VM was more beneficial than PM. For Test C, results are a bit more complicated. First of all, students in the VPP and PPP conditions showed fewer correct conceptions and more NSACs than students in the other conditions on Posttest C. Further, students in the VVV condition showed better qualitatively determined knowledge than students in the PPV and PVP conditions on Posttest C. If we compare these final two conditions, students in the PVP condition displayed more correct conceptions and fewer NSACs than students in the PPV condition, which is the only case where PM outperforms VM.

Finally, on the CEC test, the students in the VVV and PVP conditions shifted from NSACs to the SACS to a greater extent than did those in the PPV, VPP, and PPP conditions, similar to what was seen for Test B. This again is a result that is very much in line with what was found in the quantitative analysis.

Conceptual Models of Electric Circuits

Our second qualitative analysis of the test data revealed that the students in all conditions share the same scientifically incorrect, current-flow-based conceptual models, namely Conceptual Models 1 and 2, when they entered the study (see Figure 3). The most prevalent across all conditions was Conceptual Model 1, accounting for about 80% of the students in all conditions. After the students were introduced to Part A, most of the students transformed the conceptual model that they had developed before this study into a new conceptual model, namely Conceptual Model A (for frequencies, see Figure 4). This particular model was appropriate for predicting and explaining the behavior of one- and two-bulb circuits (how one- and two-bulb circuits function). A smaller number of students failed to develop Conceptual Model A and developed an alternative to it
<table>
<thead>
<tr>
<th>Conception Type</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAC</td>
<td>1.9 (0.5)</td>
<td>4.2 (0.5)</td>
<td>1.8 (0.4)</td>
<td>4.3 (0.5)</td>
<td>1.9 (0.4)</td>
<td>4.4 (0.5)</td>
<td>2.0 (0.4)</td>
<td>4.4 (0.6)</td>
<td>1.9 (0.4)</td>
<td>4.3 (0.6)</td>
</tr>
<tr>
<td>NSAC</td>
<td>5.5 (1.0)</td>
<td>2.5 (0.9)</td>
<td>5.7 (1.1)</td>
<td>2.3 (0.8)</td>
<td>5.7 (0.6)</td>
<td>2.4 (0.7)</td>
<td>5.6 (0.5)</td>
<td>2.7 (0.8)</td>
<td>5.4 (0.7)</td>
<td>2.9 (0.7)</td>
</tr>
<tr>
<td>SAC</td>
<td>0.9 (0.5)</td>
<td>3.6 (0.7)</td>
<td>0.9 (0.5)</td>
<td>4.3 (0.7)</td>
<td>1.0 (0.6)</td>
<td>3.7 (0.8)</td>
<td>1.0 (0.6)</td>
<td>4.2 (0.7)</td>
<td>1.0 (0.5)</td>
<td>3.7 (0.6)</td>
</tr>
<tr>
<td>NSAC</td>
<td>5.8 (1.3)</td>
<td>3.3 (1.1)</td>
<td>5.5 (1.2)</td>
<td>2.5 (1.0)</td>
<td>5.9 (0.8)</td>
<td>3.3 (1.0)</td>
<td>5.9 (1.8)</td>
<td>2.6 (1.0)</td>
<td>5.7 (0.7)</td>
<td>3.4 (0.8)</td>
</tr>
<tr>
<td>SAC</td>
<td>1.2 (0.4)</td>
<td>4.1 (0.6)</td>
<td>1.2 (0.4)</td>
<td>4.8 (0.7)</td>
<td>1.3 (0.4)</td>
<td>4.2 (0.7)</td>
<td>1.3 (0.5)</td>
<td>4.7 (0.8)</td>
<td>1.3 (0.4)</td>
<td>4.4 (0.5)</td>
</tr>
<tr>
<td>NSAC</td>
<td>5.4 (1.2)</td>
<td>2.7 (1.0)</td>
<td>5.5 (1.4)</td>
<td>1.8 (0.8)</td>
<td>5.5 (0.6)</td>
<td>2.8 (0.9)</td>
<td>5.5 (0.7)</td>
<td>2.0 (1.0)</td>
<td>5.2 (0.5)</td>
<td>2.3 (0.7)</td>
</tr>
<tr>
<td>SAC</td>
<td>2.0 (0.9)</td>
<td>10.2 (1.8)</td>
<td>1.9 (0.9)</td>
<td>12.7 (1.9)</td>
<td>2.1 (0.9)</td>
<td>10.4 (1.9)</td>
<td>2.1 (1.0)</td>
<td>12.3 (2.3)</td>
<td>2.0 (0.8)</td>
<td>11.1 (1.5)</td>
</tr>
<tr>
<td>NSAC</td>
<td>13.2 (2.4)</td>
<td>4.1 (1.4)</td>
<td>13.4 (2.3)</td>
<td>2.5 (1.3)</td>
<td>13.4 (1.3)</td>
<td>4.2 (1.2)</td>
<td>13.5 (2.1)</td>
<td>2.8 (1.5)</td>
<td>13.7 (2.5)</td>
<td>3.6 (1.0)</td>
</tr>
</tbody>
</table>

Note. SAC denotes scientifically acceptable conception. NSAC denotes not scientifically acceptable conception. VVV: participants used VM alone; PVP: participants used both PM and VM by following the sequence PM-VM-PM; PPV: participants used both PM and VM by following the sequence PM-PM-VM; VPP: participants used both PM and VM by following the sequence VM-PM-PM; PPP: participants used PM alone.
FIGURE 3 The conceptual current-flow-based models students developed throughout the study. It includes the conceptual models that students had prior to this study (Conceptual Models 1 and 2); the scientifically accurate Conceptual Models A, B and C that some students managed to develop after studying Parts A, B and C, respectively; and the most prevalent Alternative Conceptual Models A, B, and C (scientifically non-accurate conceptual models) that some other students developed after studying Parts A, B, and C, respectively. Conceptual Model B builds upon Conceptual Model A and Conceptual Model C builds upon Conceptual Model B; the bold letters indicate the new concepts that are incorporated in each conceptual model. For the alternative conceptual models, the bold letters indicate how Alternative Conceptual Models A, B, and C differ from Conceptual Models A, B, and C, respectively.2

(Alternative Conceptual Model A). As shown in bold letters in the Alternative Conceptual Model A box in Figure 3, the students who developed Alternative Conceptual Model A diverted from Conceptual Model A primarily because of NSACs they had in mind concerning the battery and its relation to current (e.g., a battery is a source or repository of current, a battery always provides the same amount of current in a circuit, and what current consists of exists only in batteries) and the

2Most of the students thought that the current passing through the battery was coming out of the positive terminal, going through the circuit and then back to the negative terminal of the battery. However, there were some students (much fewer) who described the current flow as going the other way (from the negative to the positive terminal).Resistance per se and calculating total resistance were introduced in Part C. In Part B, resistance was understood in a different manner. Adding bulbs and/or networks of bulbs in series increased the obstacles resisting the current flow in an electric circuit branch or route; thus, resistance increased. Adding branches with bulb(s) or a network of bulbs in parallel increased the number of routes through which the current flow could pass, which meant that resistance decreased.
PHYSICAL AND VIRTUAL MANIPULATIVES

FIGURE 4 Students' conceptual models as the curriculum progressed. Arrows indicate the paths along which students changed their conceptual models. The numbers in parentheses give the number of students in each condition. The dashed arrow shows that some students transformed their Alternative Conceptual Model B to Conceptual Model B before adopting Conceptual Model C (in Part C).

nature and behavior of the current in a circuit (e.g., for a battery and two bulbs connected in series, the current coming out of the battery is equally shared between the two bulbs; this conflicts with the fact that charges move only in one direction and that there is only one way for the current or charges to move through such a circuit). After Part B, most of the students who used VM during Part B (PVP and VVV conditions) further developed their conceptual model (Conceptual Model B), turning it from a qualitative to a semiquantitative current model by incorporating Kirchhoff's first rule into it, whereas the students who used PM (PPP, VPP, and PPV conditions) failed to do so. The PM users created an alternative conceptual model (Alternative Conceptual Model B) in which they were predicting and explaining the behavior of multi-bulb circuits but with rules and ideas that were only applicable for one- and two-bulb circuits. As a result, several NSACs included in Alternative Conceptual Model A were retained in Alternative Conceptual Model B. For example, the NSACs concerning the battery and its relation to current (e.g., for a given battery, the magnitude of the current is always the same) were preserved. Moreover, Alternative Conceptual Model B involved NSACs concerning the behavior of the current in circuits with a battery and bulbs or a network of bulbs connected in parallel (e.g., the current flow coming out of the battery divides up equally when it reaches a junction or node at which one end of the branches of bulbs are connected together according to the number of branches). A noteworthy finding concerning the PM conditions (PPP, VPP, and PPV) was that a small number of students (most coming from the same group) did manage to develop Conceptual Model B.
After Part C, most of the VVV and PVP students further extended their current-flow-based model developed in Parts A and B into a model for electric circuits that accounted for the concepts of resistance and voltage as well as the concept of current (Conceptual Model C). The model remained semiquantitative in nature, which resulted in associating the amount of current flow with the concepts of resistance and voltage. The majority of PPV students were also found to develop Conceptual Model C, but most of them had a different starting point (Alternative Conceptual Model B) than the VVV and PVP students (Conceptual Model B). As a result, the PPV students were found to develop Conceptual Model B in Part C prior to developing Conceptual Model C. This implies that the presence of VM in Part C still managed to benefit PPV students’ understanding of concepts introduced in Part B; it worked like a recovery mechanism. On the other hand, the PPP and VPP students further extended their current-flow-based Alternative Conceptual Model B and transformed it into another alternative conceptual model (Alternative Conceptual Model C), which incorrectly accounted for aspects of the concepts of voltage and resistance as well. In this new Alternative Conceptual Model C, more NSACs were added, which were combined with the NSACs already included in students’ Alternative Conceptual Model B. In particular, these students further enriched their model with NSACs concerning the battery and its relation to current (e.g., the more batteries I add in series, the bigger current source or repository I have; the more batteries I add in series, the more current will be flowing in the circuit, regardless of whether the total resistance of the circuit changes). Additionally, in Alternative Conceptual Model C students were still found to have an incorrect understanding about the behavior of the current in circuits with batteries and bulbs or networks of bulbs connected in parallel (e.g., the current entering a network of bulbs connected in parallel divides up equally according to the number of branches, regardless whether each branch might have different total resistance).

An interesting finding concerning the PPP and VPP conditions was that the students in these conditions who developed Conceptual Model B also developed Conceptual Model C. Remarkably, a few more PPP and VPP students developed Conceptual Model C after Part C, even though right after Part B they had developed an alternative conceptual model (Alternative Conceptual Model B). As with the PPV students above, these PPP and VPP students were found to develop Conceptual Model B in Part C prior to developing Conceptual Model C. Finally, some of the students in the PPP and VVP conditions who developed an alternative conceptual model after Part B (Alternative Conceptual Model B) also developed Conceptual Model C by the end of the study.

There are several questions to be raised at this point. Why could no differences between conditions be found with regard to the conceptual model developed right after Part A, whereas differences could be found after Parts B and C? What caused the majority of PPP and VPP students to deviate from developing scientifically correct conceptual models after Part A as the VVV and PVP students did? Does the additional affordance carried by VM (i.e., a view of the charge flow) have anything to do with the fact that differences in conceptual knowledge could be found among the study’s conditions after Parts B and C? If yes, what does this mean about the fact that PVP students used PM in Part C and no differences were found between the PVP and VVV conditions after Part C?

These are questions that cannot be answered by using the test data collected. Below we analyze data that relate to the processes in which students engaged during PM or VM experimentation, in an attempt to understand the circumstances under which PM and VM experimentation began to differentiate and have a different effect on students’ conceptual understanding.
Problems Encountered and Their Effect on the Processes Followed During Experimentation and Students’ Conceptual Understandings

In our effort to understand the reasons students’ conceptual understanding was found to develop differentially between conditions, we examined whether students in different conditions encountered different problems during experimentation and whether these problems differentially affected their experimentation processes and subsequently their conceptual understanding. In so doing, we used the data collected through video and audio recordings and the instructors’ journals. We also used the interview data for clarification purposes.

Problems Encountered. The analysis of the instructors’ journals data revealed three major categories of problems that students encountered during experimentation. The first category involved problems (process-related in nature) that concerned the feedback received from the manipulatives used, particularly from PM (e.g., PM did not provide observable feedback in some circuits). The second category involved problems, again process-related, that concerned the construction of complex circuits (e.g., students using PM faced problems with constructing circuits that included networks of bulbs—branches of bulbs connected in parallel—that were connected in series). The third category concerned conceptual problems that were caused by the NSACs that students held about certain electric circuit concepts.

The chi-square goodness-of-fit test revealed no statistically significant differences between the groups (students worked in groups of two or three) from each condition separately in terms of these problems (electric circuit feedback, building complex circuits, and conceptual) that the students encountered during experimentation. The chi-square value was found to be low and in all cases it was not significant. This implies that the groups within a condition faced about the same problems during experimentation.

The chi-square test (Cramer’s $V$ statistic) revealed that among the five conditions, there were significant differences in terms of the problems encountered while conducting certain experiments. In particular, it was found that the five conditions differed in 13 out of the study’s 50 experiments. Nine out of these 13 experiments were included in Part B of the study. The remaining four experiments came from the Part C of the curriculum, and all of them concerned Section 8, which covered Kirchhoff’s second rule. Cramer’s $V$ was found to be higher than 0.7 (at the $p < .001$ level; the $p$-value was set according to familywise error techniques [Hochberg & Tamhane, 1987], $p < .05/50 = .001$, to avoid the possibility of errors) for all of these experiments.

In order to identify exactly which condition(s) differed from the others, we proceeded with post hoc pairwise comparisons (10 pairwise comparisons per experiment; e.g., PPP vs. VVV, PPP vs. VPP, VPP vs. PVP) by using the same statistical test (chi-square and Cramer’s $V$). It was found that in the nine experiments of Part B, the PPP, VPP, and PPV conditions differed significantly from VVV and PVP conditions (Cramer’s $V$ was higher than 0.74 in all cases at the $p < .0003$ level; the $p$ value was set according to the family-wise error techniques of Hochberg and Tamhane [1987], $p < .05/130 = .00038$). No significant differences were found between the VVV and PVP conditions and none between the PPP, VPP, and PPV conditions. This latter finding implies that in Part B, the cause of differentiation between the conditions was the presence of VM in the VVV and PVP conditions. Moreover, a close look at the cross-tabulation statistics for these analyses revealed that the majority of the groups from the PPP, VPP,
and PPV conditions faced problems from all three categories (electric circuit feedback, building complex circuits, and conceptual) in these nine experiments of Part B, whereas the majority of the groups from the VVV and PVP conditions faced only concept-related problems. This means that the source of differentiation was the presence of process-related problems (electric circuit feedback related problems and building complex circuits related problems) for students in the PPP, VPP, and PPV conditions. This implies that these problems originated from the use of PM in Part B.

For the four experiments identified as differentially problematic in Part C, a different pattern was observed than in Part B. More specifically, the PPP and VPP conditions were now found to differ significantly from the VVV and PPV conditions, and the PVP condition was found to differ significantly from all four other conditions (Cramer’s $\hat{V}$ was higher than 0.74 in all cases at the $p < .001$ level; the $p$ value was set according to the family-wise error techniques of Hochberg and Tamhane (1987), $p < .05/40 = .0012$). No significant differences were found between the VVV and PPV conditions and none between the PPP and VPP conditions.

A close examination of the cross-tabulation statistics for these analyses revealed that the majority of the groups in the PPP, VPP, and PVP conditions faced problems from all three categories (electric circuit feedback, building complex circuits, and conceptual) in these four experiments of Part C. However, the PVP student groups handled the process-related problems encountered on their own, whereas the PPP and VPP groups requested support from their instructors to handle process-related problems. The majority of the groups in the VVV and PPV conditions faced only conceptual problems. This means that the cause of differentiation was again the fact that only the students in the PPP, PVP, and VPP conditions faced process problems, related to both electric circuit feedback and building complex circuits. We can reasonably argue that these process-related problems are PM dependent, but we should also highlight the fact that PVP students handled such problems on their own, even though they used PM in Part C.

In an attempt to explain why PVP differed from the PPP and VPP conditions, because it also involved the use of PM in Part C, we took a closer look at the four experiments of Part C. In so doing, we found that the circuits involved in these four experiments were similar in complexity to the circuits that the PVP students experienced in Part B, when they were using VM. Hence, the fact that PVP students experienced complex circuits through the use of VM before entering Part C may have provided them with the knowledge (conceptual model) required for handling Part C’s circuits (although the circuits in Part C differed in content from those presented in Part B), despite the fact that they had switched to PM in Part C and encountered process-related problems. It appears that the experience with VM and complex circuits is carried along, even if they have to switch to PM use. In particular, if we take into consideration the conceptual models analysis, it could reasonably be argued that the current-flow-based model developed by PVP students in Part B and the knowledge of how complex electric circuits are built and function was carried along and used in Part C, which ultimately enabled students to handle process-related problems when complex circuits were introduced on their own. This finding is particularly important because it provides us with a likely explanation as to why students in the PVP condition performed differently on the CEC posttest and Posttests B and C than did those in the PPP and VPP conditions. This finding could also explain why the PVP and VVV conditions had a similar effect on students’ performance on the study’s tests.

However, the question about why the VVV and PVP conditions did not encounter any process-related problems in Part B in the first place still remains. The basic difference between the VVV
and PVP conditions and the rest of the conditions (PPP, VPP, PPV) was that the students in these two different clusters of conditions were using different manipulatives. As reported above, the students using VM in Part B did not encounter any process-related problems. Apparently, VM offered students something that PM did not. If we also take into consideration that knowledge acquisition in the *Physics by Inquiry* curriculum is evidence-based (conceptual understanding is built upon evidence and observations), it becomes obvious that the reason for the difference was that VM supplied students with better evidence and observations than PM. Here is where the unique affordance of the provision of the view of charge flow by the VM environment must have offered an advantage over PM, namely that VM always gave students observable feedback.

Overall, the instructors’ journal data analysis of the problems that the students encountered during experimentation showed that the use of PM and VM in any of the study’s conditions were both associated with conceptual problems. In fact, conceptual problems were found in most of the experiments of the study’s curriculum in all conditions. According to our qualitative analysis of the test data, these conceptual problems were caused by the same conceptions that were not SACs that students in all conditions held, more or less. On the other hand, this analysis revealed that only the use of PM was associated with process-related problems when complex circuits were introduced and that only students in the PVP condition handled these problems on their own in Part C after studying complex circuits with the use of VM in Part B. It appears that the conceptual current-flow-based model of complex electric circuits developed in Part B, when PVP students were using VM with its view of the charge flow, was carried along to Part C, where the students were using PM. Of course, this does not imply that prior use of VM always enhances subsequent PM use. If this was true, then the students in the VPP condition would not encounter any process-related problems in either Part B or Part C. Apparently, it is not always enough to have VM preceding PM. The point in the curriculum at which VM should precede PM use is another crucial factor. In our case, it was found that VM needed to be used when the complexity of electric circuits started to increase, and process-related problems started to emerge for PM users for the first time (the beginning of Part B). In a way, having VM precede PM when complex circuits are introduced works as a fading scaffold. Using first VM and then PM for complex circuits introduces the difficulties one at a time; first the conceptual difficulties are introduced (during VM use) and then the procedural ones are introduced (during PM use). Doing it the opposite way (first PM and then VM) confronts students with both conceptual and procedural difficulties at the same time, which could interfere with sense making and learning. These findings are particularly important because along with identifying the process-related problems that affect students’ learning, which were more or less expected when using PM, they provide contextuality. These findings offer the particulars of where critical learning is going on in circuits and of when circuits pass the point of easy inspectability, and the use of VM appears to be advantageous.

For triangulation purposes, we checked the interview data analysis and found that the interviewees who used PM reported problems falling under the same two categories of process-related problems as found through the instructors’ journals data analysis, namely building complex circuits and problematic feedback. Specifically, all the interviewees who used PM referred to these problems, including the PVP interviewees. None of the interviewees who used VM reported any process-related problem concerning VM use. These findings validate what was found in our instructor journals’ data analysis. Below we provide quotes from interviewees who were in the
conditions that included the use of a combination of PM and VM. As shown in the excerpts below, these students highlight these process-related problems that are associated with PM use and make reference to the fact that they did not encounter such problems when using VM. Specifically, they refer to the fact that VM always offered them feedback, either through directly observing the bulbs or through the view of the charge flow.

Interviewee 3 (PPV condition): I liked the virtual lab more than the physical material because it provided precise feedback and it was much easier to build big [complex] circuits with. . . . At no point during the use of the virtual lab did we get into a dispute about a circuit or an observation. . . . With physical material we had the opposite. Whenever we did not have a clear observation we argued about it, which always ended up in rebuilding the circuit. . . . This was another problem of the physical [PM]. . . . Rebuilding the circuit never resolved our disputes. The doubt was always there.

Interviewer: What doubt?
Interviewee 3: Whether the circuit was built correctly, whether the bulbs were fine, whether the battery was dead, whether all elements were connected to each other.

Interviewee 5 (PVP condition): I liked both [PM and VM]. However, if I had to choose, I would pick the virtual lab.

Interviewer: Why?
Interviewee 5: With physical material we had problems with our observations in some of the experiments after Section 6 [Part C]. . . . With the virtual lab we did not face such problems. We were always able to make an observation and to confirm it through the charges view tool.

Interviewer: What do you mean you had problems with your observations in some of the experiments after Section 6?
Interviewee 5: For example, for an unlit bulb we could not conclude whether no current was passing through it or whether the amount of current passing through it was small.

Interviewer: What did you do when that occurred in order to reach a conclusion?
Interviewee 5: We used what we learned in previous sections. . . . Some of the experiments were the same as before [there was a small overlap between the circuits of each part of the curriculum for comparison purposes; the circuits of the four aforementioned experiments, in which differences were found among conditions due to process-related problems, were not the same]. We also used the notes we took from before [notes taken by students in their lab notebook when conducting experiments for previous sections; PVP students could find information about predicting and explaining the behavior of complex circuits in these notes, since they had experienced complex circuits in Part B].

Interviewer: How did your notes help you out?
Interviewee 5: We had notes on how current flows in different multibulb circuits, which enabled us to conclude how current flows when networks of bulbs are in series or in parallel.

Interviewer: Can you be more specific?

---

3Each excerpt provided follows the same transcription conventions: (a) the text in brackets (e.g., [PM]) clarifies what the interviewees were referring to, (b) the three periods in a series denotes ellipsis, which is an intentional omission of details (i.e., a word, a sentence, or a whole section from text) from the excerpt without altering its original meaning, and (c) the three periods in a bracket ([ . . . ]) denote that part of the conversation is intentionally omitted. The excerpt of each interviewee concerns a separate conversation.
Interviewee 5: Yes, we concluded that when bulbs or networks of bulbs were connected in series the amount of current flow got less. When bulbs or networks of bulbs were connected in parallel the amount of current increased.
Interviewer: What helped you to reach these conclusions?
Interviewee 5: The sections in which we used the virtual lab.
Interviewer: Which sections are you referring to?
Interviewee 5: Sections 3, 4 and 5 [Part B].
Interviewer: Why these sections?
Interviewee 5: Those were the ones that had the similar circuits [the student was referring to Section 8 in Part C, which also included complex circuits].
Interviewer: Why didn’t you face the same problems as the ones you faced after Section 6, in Sections 3, 4 and 5?
Interviewee 5: We had the virtual lab for these sections. . . . We always were able to make an observation.

The PPP and VPP students also expressed similar statements about the process-related problems encountered. They also painted a negative picture of PM in the experiments where no clear feedback was available or complex circuits had to be built, which also led to sense of mistrust about the validity of a circuit in an experiment (this was also evident in the video data analysis; see the Problems Encountered and Processes Engaged in during Experimentation section below). The following excerpt is particularly revealing about the mistrust that students developed towards the use of PM when building electric circuits.

Interviewee 23 (VPP condition): In certain experiments we could not tell whether the circuit was built correctly or not. We had certain expectations but we could not figure out whether these expectations were right or wrong because most of the bulbs did not glow. Was it because we built the circuit incorrectly? Was it because a bulb was broken or shorted out? Was it because the circuit was open because the circuit elements were not appropriately connected? Was the battery dying? It was a mess that we could not sort out even after checking the circuit several times or even rebuilding it . . . it got to the point where we could not trust the circuit and we had to talk to the instructor.

Another important aspect that emerged from the interview data was how the PVP students managed to handle process-related problems on their own when complex circuits were introduced in Part C. The following excerpt is representative of the PVP interviewees and is particularly supportive in terms of our hypothesis that PVP students used the predictive power that the current-flow-based conceptual model they developed in Part B gave them for complex circuits to handle the process-related problems that emerged from the use of PM in Part C.

Interviewee 6 (PVP condition): . . . we felt that our observations were problematic in some of the experiments of Section 8 [Part C] . . .
Interviewer: What do you mean by problematic observations?
Interviewee 6: Some bulbs were not lighting. We did not know whether current was passing through them or not and we could not decide about the voltage across these bulbs.
Interviewer: What did you do to reach a conclusion about these bulbs?
Interviewee 6: We used what we knew from before.
Interviewer: What exactly did you use?
Interviewee 6: What we learned from previous sections.
Interviewer: Which sections?
Interviewee 6: Sections 3 and 4 [Part B].
Interviewer: How about Section 5, which was on Kirchoff’s first rule [also in Part B]?
Interviewee 6: Yes, that one too.
Interviewer: What exactly helped you from these sections to reach conclusions for circuits that offered problematic observations in Section 8, as you said before?
Interviewee 6: We applied what we had learned about how current flows in circuits with many bulbs and networks of bulbs [complex circuits]. Even though they were not the same circuits and new things were added, such as more batteries, we still could infer how current flows in a new circuit based on the rules we wrote down [in their lab notebook] in previous sections [Part B] about how current moves in a circuit with many bulbs and in different configurations. . . . These rules helped us predict if a bulb has current passing through it and how this current compares to other bulbs’ current . . .

Interviewer: How did you decide what to follow? I am referring to observations versus prior knowledge and rules concerning current flow.
Interviewee 6: When the observations were not clear and there were disagreements among us, we used knowledge and rules about current from before. This also helped in convincing each other and then proceeding with the experiment.
Interviewer: How confident were you about the correctness of this knowledge and the rules about current?
Interviewee 6: Very confident. They worked for all circuits for which we had clear observations.

Interviewer: When did you have clear observations?
Interviewee 6: When the circuits had a small number of bulbs and for more complex circuits with more bulbs and networks of bulbs, only when using the virtual lab [Part B].

Interviewer: Did you always have clear observations with the virtual lab?
Interviewee 6: Yes. Through the brightness of the bulbs and through the charge view.

Interviewer: How often did you use the charge view in the virtual lab?
Interviewee 6: Very often. It helped a lot.

Interviewer: How?
Interviewee 6: You could tell for sure if current was passing through the bulb or whether it was shorted. . . . From how fast the charge was moving you could also tell about the amount of current passing through a bulb over time and about the voltage of the battery and the resistances . . .

[. . .]

Interviewer: Coming back to your knowledge and rules about current, how exactly did they help you reach conclusions about new concepts, such as the voltage across each bulb?

Interviewee 6: We could tell from the current passing through them about their voltage. For identical bulbs, the more current passing through it, the more voltage it had . . .

Finally, the interview data analysis revealed that the interviewees, across all conditions, did not mention anything problematic about the curriculum’s instructional methods (e.g., working in groups, no lecturing) or the curriculum used. There were also no particular problems reported as far as the sequence of experimental modes the students experienced. Only three out of the five PVP students reported that they did not like the fact that they had to switch from VM to PM when entering Part C. They felt that the VM better served their needs.
Problems Encountered and Processes Engaged in During Experimentation

Given that the students in the five conditions were found to differ in the process-related problems they encountered in the particular 13 experiments mentioned, we purposefully selected the video data for these 13 experiments (for all conditions) to examine whether these process-related problems affected student processes during experimentation. In particular, because the problems were encountered only when students used PM, we examined how PM experimentation processes differed from VM ones. We used students’ discourse and actions as means to identify the processes in which students engaged during PM and VM experimentation. We isolated all the episodes in these 13 experiments (critical events), in which process-related problems emerged and analyzed students’ discourse and actions. This was an attempt to further refine our understanding of how these differentially problematic points during experimentation may affect students’ conceptual understanding.

The analysis revealed 15 different codings reflecting students’ discourse and actions across all episodes (see Figure 5). The timeline graphs produced after the isolation of the critical events revealed similar interrelationships and patterns of the codings (student talk and actions) over time between the conditions that faced process-related problems and between the conditions that faced process-related problems and those who did not.

Specifically, when the students in the PPP, VPP, and PPV conditions encountered problematic feedback (when no observable outcomes came out of a circuit) in Part B, it was found that they first engaged in a discussion disputing whether they had built the correct circuit and then discussed rebuilding the circuit from the beginning (for a representative example, see Graph 5a of Figure 5). This pattern was observed to occur more than once in some cases. After this iterative pattern, these students engaged in a discussion with an instructor, seeking support and expressing their mistrust of the feedback received from the circuit (see Lines 4–8 in Graph 5a of Figure 5). On the other hand, the students who did not encounter this process-related problem (students in the VVV, PVP [only in Part B], and PPV [only in Part C] conditions), were never involved in such discourse and action patterns because they always had observable feedback (if not through the bulb’s brightness, then through the view of the charge move; for a representative example, see Graph 5b of Figure 5). Interestingly, similar discourse and actions patterns as those observed in the case of VVV were shown for the PVP condition, even though the PVP students were using PM in Part C and encountered process-related problems. The only difference in terms of discourse and action patterns between the PVP and VVV conditions in Part C (in Part B they followed about the same discourse and action patterns) was that the PVP students discussed the validity of the circuit built and their observations (Code 7 of the graphs of Figure 5). But the PVP students almost never proceeded with rebuilding a circuit and never asked for help from the instructors as the PPP and VPP students did. This finding agrees with the conclusions that emerged from the conceptual test analysis, which indicated that the conceptual model developed about complex electric circuits in Part B, when PVP students were using VM, was transferred to Part C and enabled them to handle the process-related problems encountered in Part C on their own (for a representative example, see Graph 5d of Figure 5). Comparing the PPP and VPP patterns to those for PVP for Part C, there is no doubt that the presence of VM in Part B was the factor that caused the difference between them.

Moreover, it is important to note that the students using PM (with the exception of the PVP condition in Part C) spent most of their time on rebuilding a circuit and disputing its validity...
FIGURE 5 Representative examples of student discourse and actions. Graph 5a presents the discourse and actions over time of PPV students (Group 1 of the PPV condition) using PM to conduct experiment 3.3B (from Part B of the curriculum). Graph 5b presents the discourse and actions over time of PVP students (Group 2 of the PVP condition) using VM to conduct experiment 3.3B (from Part B of the curriculum). Graph 5c presents the discourse and actions over time of PPV students (Group 2 of the PPV condition) using PM to conduct experiment 3.6A (from Part B of the curriculum). Graph 5d presents the discourse and actions over time of PVP students (Group 1 of the PVP condition) using PM to conduct experiment 8.12A (from Part C of the curriculum) (Continued).


**Student discourse and actions**

1. Students read the instructions and/or the accompanying questions (as provided in the teaching material)
2. Students make predictions (as instructed by the teaching material)
3. Students discuss their predictions (any prediction is accompanied with the reasoning behind it)
4. Students (re)build the circuit of the experiment (according to the teaching material)
5. Students make observations
6. Students discuss their observations
7. Students discuss the validity of the circuit built and their observations
8. Students call their instructor and discuss with her the validity of the circuit built and their observations (this is not suggested by the curriculum as a checkpoint. The students themselves requested a discussion with their instructor).
9. Students write down their observation and explain if they differ from their predictions (as instructed by the teaching material)
10. Students discuss the accompanying question and its possible answer
11. Students write down their answer in their lab notebook
12. Discussion on an issue that is irrelevant to the class
13. Students further experiment with the circuit built by adding or subtracting more bulbs (in this case they added a third bulb in parallel with bulb B and then a fourth one)
14. Students discuss how to build certain parts of the circuit (the student discussion is accompanied with corresponding demonstrations)
15. Students call their instructor and discuss with her how to build certain parts of the circuit

*FIGURE 5* (Continued)

(about 30% to 50% of the time spent on an experiment; for representative examples, see Graphs 5a and 5c of Figure 5), rather than on repeating an experiment to resolve conflicts or questions raised by group members during an experiment, as students using VM did (see Graph 5b of Figure 5). In other words, students using VM engaged in constructive actions and discussions, which was based on the students’ trust of the circuits built, whereas students using PM (with the exception of the PVP condition in Part C) engaged in procedural actions and discussions that resulted from students’ problematic attitude toward the circuits built.

For the building complex circuits problem category, it was found that the students using PM (with the exception of the PVP condition in Part C) discussed among themselves how to build certain parts of a circuit and then complete the whole circuit. Then in checking the outcomes of the circuit (bulbs’ brightness), they discussed among themselves about whether the circuit was built correctly (see Graph 5c of Figure 5). Whenever the students of a group using PM disagreed about whether the circuit was built correctly, they engaged in a discussion with an instructor (for an example, see Line 8 in any of the graphs of Figure 5). As in the case of the problematic feedback category, this resulted in the PM students spending more time on procedural issues (with the exception of the PVP condition in Part C) rather than on actions and discussions that were more constructive for learning, as students using VM did. The time issue worsened for PM students when the building complex circuits and the problematic feedback problems co-occurred (see Graph 5c as opposed to Graph 5a of Figure 5). In this case, the action pattern for the students transformed into the following: (1) discussing in their group about how to build certain parts of a circuit and then complete the whole circuit, (2) checking the outcomes of the circuit (bulbs’
problems encountered while setting up and conducting an experiment led them to enact specific conversations and action patterns that proved to spend valuable learning time on procedural tasks, rather than on constructive actions and discussions. Only the PVP students were found not to follow this problematic pattern, most probably because of the conceptual current-flow-based model they developed when using VM in Part B. Additionally, it was found that the constant exposure of students using PM to circuits that did not provide observable feedback created a sense of mistrust towards the validity of an experiment’s circuit. These findings imply that the use of VM should be preferred over PM when such problems are likely in a learning process, such as experimentation in which observation has a vital role. The interview data are particularly revealing in this respect. For instance, the following interviewee stated:

Interviewee 2 (PPV condition): The virtual lab [VM] feedback was quite different from the feedback that was provided through the concrete electric circuits [PM]. First, its [VM] feedback was always precise and consistent, as well as observable. For example, I did not have to speculate or think whether a bulb had current passing through it when it was not lit . . . what is good about the virtual lab is that it provides different and discrete levels of brightness and also shows the current flow. . . . If I were using the virtual lab [rather than using the physical electric circuits] when we tried to develop the model for electric current, I believe that I would be able to concentrate mostly on building the current model [conceptual model that enables students to predict and explain the behavior of electric circuits] instead of anything else . . .

Problems Encountered and Conceptual Understanding

Because we found the five conditions to differ only in terms of process-related problems in the particular 13 experiments identified, we proceeded with a quantitative analysis to examine whether the presence of these process-related problems affected students’ level of understanding of the electric circuit concepts introduced through these 13 experiments. Specifically, we used the data collected through the instructors’ reflective journals (the process problems encountered by the students and the mean score for the level of understanding calculated for each group per experiment). Before running the two-way ANOVA, we transformed the data by assigning to each group of a condition the dominant code of the process-related problems faced in the nine experiments of Part B and by calculating a total understanding score from the scores assigned to each group per experiment for all nine experiments. The same data transformation was followed for the four experiments of Part C.

For the nine experiments of Part B we found a main effect of condition, $F(4, 57) = 3.6, p < .01$, partial $\eta^2 = .21$, and of the process-related problems encountered on students’ conceptual
TABLE 10
The Dominant Process-Related Problems Faced by the Groups in Each Condition and the Corresponding Mean Scores (and Standard Deviations) of Students’ Conceptual Understanding Scores for the Nine Experiments of Part B for Which the Conditions Differ

<table>
<thead>
<tr>
<th>Condition</th>
<th>Electric Circuit Feedback/Building Complex Circuits (Mean, SD)</th>
<th>Electric Circuit Feedback Alone (Mean, SD)</th>
<th>Building Complex Circuits Alone (Mean, SD)</th>
<th>None Encountered (Mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPP</td>
<td>11 (9.7, 1.0)</td>
<td></td>
<td></td>
<td>2 (13.5, 0.7)</td>
</tr>
<tr>
<td>VVV</td>
<td>—</td>
<td></td>
<td></td>
<td>13 (20.8, 1.7)</td>
</tr>
<tr>
<td>VPP</td>
<td>13 (10.1, 1.0)</td>
<td></td>
<td>1 (12.0)</td>
<td></td>
</tr>
<tr>
<td>PVP</td>
<td>—</td>
<td></td>
<td>1 (19.0)</td>
<td>12 (21.0, 1.6)</td>
</tr>
<tr>
<td>PPV</td>
<td>12 (9.8, 1.0)</td>
<td></td>
<td>1 (13.0)</td>
<td></td>
</tr>
</tbody>
</table>

Note. *The number preceding the parentheses. All conditions have 13 groups of students except the VPP condition, which has 14 groups.

understanding, $F(2, 57) = 10.5, p < .01$, partial $\eta^2 = .27$, but no interaction effect between condition and process-related problems, $F < 1, ns$. Bonferroni pairwise comparisons showed ($p < .01$ for all statistically significant differences) that students’ understanding scores in the VVV and PVP conditions were significantly higher than the scores of the students in the PPV, VPP, and PPP conditions (for mean scores and standard deviations, see Table 10). The pairwise comparisons did not show a significant difference between the students’ scores in the PVP and VVV conditions and between the PPP, VPP, and PPV conditions. It is important to note that these findings match the findings resulting from our earlier analysis of the conceptual knowledge tests.

The fact that a main effect of the process-related problems encountered was identified in these nine experiments of Part B implies that these problems affected students’ understanding. The descriptive statistics for this analysis show that only the groups in the PPP, VPP, and PPV conditions faced process-related problems (see Table 10). In fact, the majority of these groups were found to face problems with both electric circuit feedback and building complex circuits. It appears that these process-related problems were what that caused the PPP, VPP, and PPV groups of students to score lower than those in the VVV and PVP conditions.

For the four experiments of Part C we found a main effect of condition, $F(4, 57) = 5.7, p < .01$, partial $\eta^2 = .29$, and of the process-related problems encountered on students’ conceptual understanding, $F(3, 57) = 6.8, p < .01$, partial $\eta^2 = .26$, but no interaction effect between condition and process-related problems, $F < 1, ns$. Bonferroni pairwise comparisons showed ($p < .01$ for all statistically significant differences) that students’ understanding scores in the VVV, PVP, and PPV conditions were significantly higher than scores in the VPP and PPP conditions (for mean scores and standard deviations, see Table 11). The pairwise comparisons also showed that the VVV condition received significantly higher scores than did those in the PPV condition but did not differ from the PVP condition. Additionally, the pairwise comparisons did not show a significant difference between the students’ scores in the PVP and PPV conditions nor between the students’ scores in the PPP and VPP conditions. These findings also perfectly match the findings resulting from our conceptual tests analyses. Given that these findings come...
TABLE 11
The Dominant Process-Related Problems Faced by the Groups in Each Condition and the Corresponding Mean Scores (and Standard Deviations) of Students’ Conceptual Understanding Scores for the Four Experiments of Part C for Which the Conditions Differed

<table>
<thead>
<tr>
<th>Condition</th>
<th>Electric Circuit Feedback/Building Complex Circuits (Mean, SD)</th>
<th>Electric Circuit Feedback Alone (Mean, SD)</th>
<th>Building Complex Circuits Alone (Mean, SD)</th>
<th>None Encountered (Mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPP</td>
<td>11 (5.0, 1.0)</td>
<td>—</td>
<td>2 (8.5, 0.7)</td>
<td>—</td>
</tr>
<tr>
<td>VVV</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>13 (15.8, 1.8)</td>
</tr>
<tr>
<td>VPP</td>
<td>13 (5.2, 0.9)</td>
<td>—</td>
<td>1 (8.0)</td>
<td>—</td>
</tr>
<tr>
<td>PVP</td>
<td>12 (15.5, 1.5)</td>
<td>—</td>
<td>1 (12.0)</td>
<td>—</td>
</tr>
<tr>
<td>PPV</td>
<td>12 (14.1, 1.2)</td>
<td>—</td>
<td>1 (12.0)</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. a The number preceding the parentheses. b All conditions have 13 groups of students except the VPP condition, which has 14 groups. c The PVP students manage to overcome/handle these process-related problems on their own, whereas the PPP, VPP, and PPV students requested help from the instructors.

from a different data source, our earlier findings from the conceptual test analysis are triangulated, which further adds credence to this study’s findings.

A main effect of the process-related problems encountered was also identified for the four experiments of Part C, which also implies that these problems affected students’ understanding. The descriptive statistics for this analysis show that the PPP, PVP, and VPP conditions faced process-related problems related to both electric circuit feedback and building complex circuits. However, the PVP student groups differed from PPP and VPP student groups, even though all of these groups were using PM in part C, in that they managed to handle the process-related problems encountered on their own, whereas the PPP and VPP student groups requested help from the instructors. Again, this finding shows that the process-related problems negatively affected the PPP and VPP students’ conceptual understanding and how important the preceding use of VM in Part B was for the PVP students.

DISCUSSION AND IMPLICATIONS

In the current study we investigated how VM can be integrated within a curriculum in the domain of electric circuits in which PM normally dominate. There are quite a few instances when PM need to dominate a curriculum, such as exposing students to (a) touch sensory input (e.g., feeling repulsion in magnetism, feeling the viscosity of a liquid, feeling the increase in temperature of lit filament bulbs as time passes) and touch-related skills (e.g., how to operate a concrete apparatus), (b) error and how to handle it (competency in a domain includes knowledge about measurement errors and how to deal with them [Toth et al., 2009]; VM often display a too idealized world, leading to a limited view of experimentation [Chen, 2010]), (c) the messy nature of PM experimentation, and (d) how authentic science is done in the real world in order to understand the nature of science. Using both PM and VM is the only way to keep these unique PM
affordances and also take advantage of the numerous additional affordances that VM can add to a learning environment (e.g., reification of abstract or conceptual objects, altering the natural scale of time and space, allowing students to change variables that would be impossible or unrealistic to change in the natural world), especially if the enhancement of conceptual understanding is at issue (Zacharia et al., 2008), as it was in this study. However, research so far has not shown under what circumstances VM could be used along with PM for optimizing students’ conceptual understanding; rather, it has been restricted to just alternating VM and PM and examining which sequence is more conducive to students’ learning. Consequently, prior studies (Akpan & Andre, 2000; Jaakkola & Nurmi, 2008; Jaakkola, Nurmi, & Veermans, 2009; Toth et al., 2009; Toth, Suthers, & Lesgold, 2002; Zacharia, 2007; Zacharia et al., 2008) have failed to point out the specifics of when VM or PM are really needed in an experimentation learning process. The findings of this study were particularly revealing in this respect; they revealed that not all PM and VM combinations have the same effect on students’ conceptual understanding, even if the same sequence of manipulatives is used for different parts of a curriculum (e.g., always VM preceding PM). This study also provided contextuality—particulars of when to use VM advantageously for their learning affordances in a PM-dominant curriculum. In other words, our study showed that combining PM and VM is not a matter of which mode of experimentation should precede the other, as investigated in previous studies (e.g., Toth et al., 2012), but rather a matter of when a mode of experimentation, along with its affordances, really contributes to students’ learning. For example, as will be argued in more detail below, using VM before PM does not always have the same effect.

In Part A of the study’s curriculum, where one- and two-bulb circuits were introduced, PM and VM were found to be equally conducive to the development of students’ conceptual understanding. Apparently, when used together with the curriculum material, VM (along with the presence of the view of current or charge flow affordance, through which students could see dots moving at various speeds) did not provide critical information to VM users that PM users could not get from the curriculum material and their experience with concrete material. Given that the goal of Part A was for the students to develop a current-flow-based conceptual model, the question raised is what PM offered to students that let them understand that there is a current flow in a physical circuit and thus compensated for the information VM users received from the VM affordance that provided them with a view of the current or charge flow in a circuit.

Before answering the question just above, it should be noted that we identified through the CEC pretest that all participants entered the study with alternative current-flow-based models. These models have been consistently reported in the literature as the most prevalent, alternative or naïve current-flow-based mental models that young children and adults have about the current flow in electric circuits (e.g., Driver et al., 1985, 1994). Such models are (a) the clashing currents’ model in which current is considered to leave both of the terminals of the battery at the same time and meet at an electric component (e.g., bulb); and (b) the current is used up around the circuit model in which the electric current is thought to leave one terminal of the battery, pass through the electric components in which part of the current is used up, and, finally, what is left of the current reaches the other battery terminal. Hence, the aim of Part A was to set right these alternative models, rather than helping students to develop a model from scratch. In the case of VM users, setting right students’ alternatives models was not a difficult task because, on top of the carefully selected experiment or activity sequence of the curriculum material, students had
direct access to the VM current flow affordance through which students were able to see the charges and electrons flowing in any circuit.

In the case of PM users, students followed the same experiment sequence as VM users did but had visual and touch sensory input for making observations and collecting data that could enable them to revise their alternative current-flow-based models. The idea was to have the students collect data that show that when circuit components (wires and bulbs) are connected across a battery, we have evidence that something is happening throughout the circuit. The battery and wires become warm to the touch (naked wires were used when touching was involved; the battery became evidently warm when a wire was connected across it), and the bulbs glow. However, the experiments provided no evidence about the presence of a flow as in the case of VM. The concept of flow, for the PM users, emerged from their alternative current-flow-based models, which existed in their minds before this study. Thus, they used the concept of flow as a means of explaining what was happening in the circuit (wires getting warm and bulbs lighting). Overall, the findings of the study revealed that the PM visual and touch sensory input from each of the experiments of Part A enabled students to move towards a more scientifically accepted current-flow-based model (see Figure 3). The presence of both the visual and the touch sensory input appears to have been useful for the students in this respect, since each sensory channel provided unique input. For instance, seeing bulbs light enabled students to understand that something was happening in the bulbs, whereas touching naked wires and sensing that they are warm enabled students understand that something was happening in the wires as well. These sensory inputs were also crucial for helping students overcome their misconceptions related to their initial, alternative current-flow-based models. For example, the fact that identical bulbs, connected in series, had the same brightness enabled students to see that the current is not consumed (if this were the case, then brightness would have been decreasing from bulb to bulb). The fact that students touched a naked wire as soon as it was connected across a battery and found it to be equally warm at different points at the same time enabled them to feel that what is causing the heat is instant, continuous, and uniform at all points of the wire.

In this particular context, both the visual and the touch sensory input worked as an advantageous affordance for the students using PM. It was advantageous enough to compensate for the fact that VM students had access to a view of the charge flow. Of course, the information a student receives through observing physical bulbs and touching physical wires and the view of VM representations of the charge flow are not the same. Observing and touching cannot reveal the nature of what is flowing in a circuit (electrons or charges) or the direction of the flow (from the negative to the positive terminal of the battery). Nonetheless, for the purposes of Section 2 of the Physics by Inquiry curriculum, the visual and touch sensory input afforded was adequate to support students in developing a scientifically correct current-flow-based model, accounting for the current passing through one-, two-, and three-bulb circuits.

Another lesson learned from the findings for Part A is that the VM affordance of viewing the charge flow, in a context similar to the one of Part A, is not necessary. In prior studies concerning electric circuits, researchers provided students with the current flow affordance constantly, which need not be the case should someone wants it otherwise. In fact, in some cases the constant presence of an additional affordance, such as the view of abstract or conceptual objects, could be restrictive for students’ learning (Olympiou et al., 2013). For instance, the constant presence of the affordance of the view of the charge flow in Part A, could have restricted students from developing or inventing this charge flow (moving dots) on their own through mental modeling.
In Part B, when complex electric circuits were introduced, VM was found to be more conducive to developing students’ understanding of electric circuit concepts and how electric circuits function than PM. Our analyses revealed that the main cause of this difference was the fact that PM users were found to encounter process-related problems at certain instances, while VM users did not. These instances occurred whenever a complex circuit did not offer students observable outcomes (e.g., when single bulbs and networks of bulbs—branches of bulbs connected in parallel—were connected in series) and whenever students were asked to construct a complex circuit (e.g., when constructing circuits that included networks of bulbs that were connected in series).

In the case of the feedback problem, VM appears to have an advantage over PM because VM could always offer observable feedback to the students, either through the larger range of brightness levels that were available for a bulb or through the electrical charges view. PM could provide feedback only through the bulb’s brightness, which stopped being observable for some of the bulbs in the circuit as the electric circuit became more complex. A similar finding was reported by Finkelstein et al. (2005). Moreover, even when both PM and VM offered observable feedback, VM made the observations easier for the students when comparisons between bulbs were required for the purposes of an experiment, because VM provided a larger range of brightness levels for each bulb. Hence, the presence of more distinct brightness levels made it easier for the students to decide about the relative brightness of at least two bulbs. In other words, these findings indicate that PM and VM experimentation affect students’ understanding of electric circuits concepts differently whenever they provide different possibilities for observations. These results confirm results from studies that have shown advantages of VM over PM (e.g., Bell & Trundle, 2008; Chang et al., 2008; Huppert et al., 2002). The essentiality of offering students access to observable, detailed, and distinct feedback in an experimental procedure has also been shown in studies in which PM and VM carried similar affordances (Zacharia & Olympiou, 2011). In these studies no differences between knowledge gained after using PM or VM were found because students were offered access to observations of equal quality in both the PM and VM environment, for example, when the domain concerned springs (Triona & Klahr, 2003).

For problems with constructing a complex circuit, our analyses also revealed another presumed advantage of VM over PM, namely its higher efficiency with respect to the time used to set up an experiment and discuss the conceptually core aspects of the experiment under study. In particular, PM appeared to have a disadvantage compared to VM due to the messy nature of PM. As shown in the student discourse and action analysis, PM students spent much more time on building a complex circuit than VM students. Overall, we found that PM students’ attention shifted toward the procedural aspects of experimentation rather than focusing on understanding the concepts introduced through an experiment, as VM students did. Moreover, it was found through the interviews that this repeated presence of problems in PM experimentation created a sense of mistrust in the students toward the circuits built, thus creating a lack of confidence that appears to have been maintained throughout the building of circuits. This was also apparent in the video analysis, which shows that students spent a lot of time on discussing the validity of their observations (e.g., discussions about circuit conformity, discussions about whether current is passing through a bulb, discussions about the comparative brightness of bulbs).

In contrast, the sustained consistency of the VM feedback across the experiments appeared to transform the VM feedback into an epistemic authority that worked as a scaffold for students as
they constructed understanding, even when they switched to the use of PM in Part C of the PVP condition. It might be because the PVP students saw that the conceptual model they developed through VM use in Part B held true for Part C as well. By epistemic authority we mean the property of a source of knowledge (e.g., other people, computer-based technology) that attracts learners’ trust on specific knowledge or beliefs and has determinative influence on the formation of individuals’ knowledge or beliefs (Origgi, 2004). Trust is a cognitive notion, a set of beliefs or expectations about the commitment of the trusted source of knowledge to behave in a determinate way in a context that is relevant to us, and the source of knowledge we think trustworthy “constitutes the potential pool of cognitive authorities on which we might draw” (Wilson, 1983, p. 16). Hence, it appears that the messy nature of PM experimentation, in the case of building complex circuits, led to process-related problems that in turn created a sense of mistrust in the students for the circuits built. In turn, this sense of mistrust then caused a redirection of students’ attention from the conceptual to the procedural aspects of learning and finally restricted PM users (VPP, PPV, and PPP conditions) in their development of conceptual understanding in Part B of the curriculum as compared to the VM users (VVV and PVP conditions).

Despite the fact that we did not systematically examine the issues of trust and mistrust in this study, our findings indicate that there might be a relationship between students’ trust and conceptual understanding. Needless to say, another research design and other data are needed to investigate this topic. It would be interesting to see, however, how trust or mistrust toward the use of PM and VM affects the processes in which students engage during experimentation and their conceptual understanding. Moreover, it would be nice to see whether a sense of trust transcends process-related problems, as it appears to be the case with our PVP condition in Part C.

Overall, it appears that the presence of VM (along with its affordance) in Part B of the curriculum was necessary. It provided the students with all the support needed to avoid process-related problems and focus on the conceptual aspects of each experiment. However, it should be noted that our suggestion for using VM in Part B is contingent upon the fact that our learning goal was to enhance students’ conceptual understanding. If we had different learning goals, the suggestion could have been different. For instance, if our goal was only to enable students to understand aspects concerning the nature of physical electric circuits (e.g., difficulties in building circuits with PM or problematic observations coming out of concrete circuits) and gain meaningful insights about authentic scientific inquiry through the problems they encounter by using PM, our suggestion, obviously, would have been to use PM throughout.

In Part C, which also included complex electric circuits, the same pattern was observed as in Part B. Namely, VM was found to be more beneficial for students’ understanding of electric circuit concepts and functioning than PM, which can be explained in the same manner as before. The only exception was the PVP condition. In this case, the conceptual current-flow-based model that students developed about complex electric circuits in Part B, when students used VM along with its affordance of the view of the charge flow, was brought along to Part C, where the students used PM. It appears that this conceptual model enabled students to handle better, or maybe even ignore, process-related problems that emerged from the use of PM in Part C, by using the predictive power that the conceptual model they developed in Part B gave them for complex circuits. As a result of this conceptual model, PVP students advanced their conceptual understanding more than the PPP and VPP students did in both Parts B and C.

This is a crucial finding, because it indicates that a prior targeted use of VM might be enough for students to develop the conceptual model needed (e.g., a current-flow-based model) to be
able to surpass the process-related problems that might emerge in a subsequent use of PM in a context that involves building and studying complex circuits. Moreover, the fact that the PVP condition was found to be as conducive to students’ understanding of electric circuit concepts and functioning as the VVV condition and more conducive than the PPP condition throughout the curriculum means that the presence of VM was not necessary in Part C. This also means that the integration of VM within a PM-dominant curriculum is feasible, as long as we know which part or parts of the curriculum to target with the use of VM. In our case, this part was Part B, in which the complexity of electric circuits started to increase, process-related problems started to emerge in PM use for the first time, and the conceptual current-flow-based model that predicts and explains the behavior of complex electric circuits had to be developed. Failure to develop such a conceptual model resulted in having students not understand concepts related to complex circuits and concepts introduced through them in Parts B and C.

Overall, the findings concerning Parts B and C indicate that VM is more suitable for the experiments of the curriculum that involve complex electric circuits but is not required in the case of Part C, which leaves more room for the use of PM when more PM-dependent learning goals are at issue. As indicated earlier, there are curricula that include such PM-dependent learning goals. For example, only the messy interactions with PM teach students about the underlying complexity of doing science (e.g., building complex circuits). From this perspective, our study pointed to an advantageous combination of PM and VM, namely the PVP condition, which preserves the use of both PM and VM without compromising students’ conceptual understanding.

It should also be highlighted that the PVP condition was found to be as conducive to students’ development of conceptual understanding as the VVV condition, which was the condition that carried unique affordances throughout the study and was more conducive to students’ development of conceptual understanding than the other three conditions. This further emphasizes the value of the PVP condition. The fact that it affected students’ conceptual understanding similarly to the VVV condition further supports the idea that a targeted use of VM is feasible without compromising students’ learning. Moreover, the fact that the influence of the PVP condition was greater than that of the PPV condition on the CEC test highlights the importance for learners of having the VM support the first time complex electric circuits are introduced in the curriculum.

Despite the fact that the PPV condition was found to be less influential for students’ development of conceptual understanding on the CEC test than the PVP condition, it is important to highlight the fact that although its students used VM only in Part C, most of its students managed to move from a scientifically incorrect conceptual model to a scientifically correct conceptual model about electric circuits by the end of the study. This finding shows that there might also be an added value of using VM at a later stage than Part B, when complex circuits are involved again, if other reasons restrict earlier use of the VM.

Overall, the findings of this study offer several practical implications for researchers and educators when it comes to deciding which type of manipulatives (PM and/or VM) to use for learning through experimentation. Some of these implications are context-specific, meaning that they appear to apply for contexts similar to the one in this study (with similar electric circuit curriculum using PM and VM), whereas others are more general in nature and appear to transcend the context of this study. Starting from the more general implications, the findings of the study show that students should be offered the use of manipulatives that allow better access to observations and less time for setting up an experiment when the development of conceptual understanding is desired. In our case, among the reasons that VM use was found to
be advantageous in promoting students’ conceptual understanding were that VM always offered students access to observations and that VM had higher efficiency with respect to the time used to set up an experiment and discuss the conceptually core aspects of the experiment under study. Other studies have shown PM to be more suitable in this respect. For instance, in the Zacharia et al. (2012) study, pre-K students were found to have better access to observations with the use of PM rather than the use of VM—through both visual and touch sensory input—which resulted in PM use enhancing students’ conceptual understanding more than the use of VM. It could be that with use of different PM in this study (e.g., using light emitting diodes [LEDs], instead of regular bulbs, which have a very wide range of brightness), some of the process-related problems (at least the ones related to restricting students from having clear observations) might have been overcome, and the learning outcomes might have been different. However, further research is needed to reach solid conclusions concerning this conjecture.

One other more general implication coming out of the study’s findings is that selection of PM or VM should be according to whether they provide sustained consistency in the feedback they offer. Our data indicate that students need to build a trust-oriented relationship with the source of knowledge (PM or VM) in order to be able to focus on the conceptual aspects of the learning process rather than on issues of mistrust due to problematic or doubtful PM or VM feedback.

Another general implication is that PM and VM can coexist in a science experimentation context and can be combined effectively and used to promote students’ conceptual understanding. Similar findings have been reported in previous studies both in the subject domain of electric circuits (e.g., Jaakkola & Nurmi, 2008; Jaakkola et al., 2011; Zacharia, 2007), and also in other subject domains (e.g., Akpan & Andre, 2000; Zacharia et al., 2008). However, an important aspect that differentiates this study from previous work is that this study shows that not all PM and VM combinations are equally conducive to learning, even if the same sequence of manipulatives is used for different parts of a curriculum (e.g., always VM preceding PM), which again has a practical implication in terms of when we should use PM or VM.

This latter issue is one of the major challenges that this research domain must address. In particular, future research efforts need to focus on how PM and VM should be combined to optimize learning according to the learning goals set. In prior research, PM and VM were not combined according to a framework that focused on optimizing learning, but based on the methodological needs of the study (e.g., providing equal opportunities for both PM and VM, or whether students using PM could switch to using VM and vice versa; for examples, see Gire et al., 2010; Jaakkola & Nurmi, 2008; Jaakkola et al., 2010; Toth et al., 2009, 2012; Winn et al., 2006; Zacharia, 2007; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011; Zacharia et al., 2008). As a result, there is not enough information coming from this research domain to support more well-thought-out PM and VM combinations. The implications related to this issue appear to be rather context-specific, which suggests that a future framework on combining VM and PM in a particular subject domain should have rather detailed context-dependent input. For example, in this study, it was found that VM should be implemented in a PM-dominant curriculum when the complexity of electric circuits starts to increase, process-related problems start to emerge in PM use for the first time, and the conceptual current-flow-based model that predicts and explains the behavior of complex electric circuits needs to be developed. These findings are particularly important because along with the process-related problems identified, which were more or less expected because of the presence of PM, they provide contextuality—the particulars of where critical learning occurs in circuits, when circuits pass the point of easy inspectability, and when
the use of VM appears to be advantageous. Practically, this means that educators teaching electric circuits should provide students with VM when the complexity of electric circuits increases and process-related problems start to emerge for the first time (e.g., students do not have access to clear feedback). In this way, having VM precede PM would work as a fading scaffold. Using first VM and then PM for complex circuits would introduce the difficulties one at a time: first the conceptual difficulties (during VM use) and then the procedural ones (during PM use). Doing it the opposite way (first PM and then VM) would confront students with both conceptual and procedural difficulties simultaneously, which could interfere with sense making and learning. Of course, this makes the practical aspects of using VM and PM in class rather complex, but if identifying the critical points in a curriculum is the only way to optimize our students’ learning, we need to turn the research efforts of the domain towards this direction in order to inform educators about the particulars of each subject domain. One final related practical implication is that any researcher who is aiming at developing a framework for combining PM and VM needs to consider context-specific information.

In the introduction, we discussed several studies that showed that a conjunction of PM and VM works best (Akpan & Andre, 2000; Jaakkola & Nurmi, 2008; Jaakkola et al., 2009; Toth et al., 2002, 2009; Zacharia, 2007; Zacharia et al., 2008). However, none of the earlier studies on combining PM and VM presented a three-part design in a PM-dominant context with a balanced crossing of conditions (i.e., VPP, PVP, and PPV). This is important because our study shows that not all combinations have the same impact on students’ conceptual understanding and reveals which combination could retain the use of PM and VM without compromising students’ conceptual understanding.

Future research should try to overcome the limitations of this study. First, we need studies with larger samples of the same population. In our case, it could be considered as a limitation that we had to draw our sample from two different semesters. Second, we need studies that add further detail to the findings of this study by identifying, for different subject matter and domains, the instances in which VM could be used to scaffold PM experimentation. For instance, it is important to know the minimum range of VM exposure that is needed for students’ learning to be supported in a PM-dominant context. In our case, it might be that the presence of VM was not necessary throughout Part B, or that it could have been better if VM were used only in the 13 experiments in which differences in the student discourse and actions were found. Third, it is necessary to examine what insights students derive about scientific inquiry from the presence of both PM and VM. For example, what meaningful insights about authentic scientific inquiry do students develop from the problems they encounter by using PM? Furthermore, this type of research should be extended across K–12. It would be interesting to see if these findings also hold true for primary and secondary school students.

In conclusion, the fact that VM was found to support students’ understanding of electric circuit concepts and functioning in a PM-dominant context challenges the already established norms which support using only PM for experimentation purposes in the science classroom (National Science Teachers Association, 2007). The National Research Council (2006) in one of its documents states clearly, “While reading about science, using computer simulations, and observing teacher demonstrations may be valuable, they are not a substitute for laboratory investigations by students” (p. 3). Specifically, the effectiveness of VM challenges laboratory experimentation as we traditionally experienced it through PM in a way that calls for its redefinition and restructuring in order to include VM. For instance, under some conditions tested in our study (e.g., when
students were confronted with more complex circuits), VM was found to provide scaffolding for experimentation and information about processes that might otherwise be lost in students’ struggles to make PM work. On the other hand, the advantages of VM, as well as the disadvantages that PM inherently carry within them (e.g., messy nature of science, measurement errors, no access to observations), must not disorient us from enacting authentic science and living these struggles. Students must experience authentic science (e.g., make actual circuits with PM) because only authentic science involves manipulation of material from the real world. Living and interacting within the physical world requires certain knowledge and skills that only PM experimentation could provide. Therefore, this call for reform creates the need for understanding the circumstances under which PM and VM can be used and coexist without compromising students’ learning. This could be achieved by extending the empirical base through research that focuses on this direction.

REFERENCES


Triona, L. M., & 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. doi:10.1207/S1532690XCI2102_02


APPENDIX

The Scoring Rubric for Subitem 12A(i)(b) of the CEC Posttest and Examples of Scored Student Responses

<table>
<thead>
<tr>
<th>Subitem</th>
<th>Correct Answer (Score)</th>
<th>Expected Reasoning (Score), [Total Score]</th>
<th>Examples of a Participant’s Posttest Response and Reasoning (Score), [Total Score, Condition, Participant Number]</th>
</tr>
</thead>
<tbody>
<tr>
<td>12A(i)(b). For the circuit below, compare the current passing through AF and GK. Explain your reasoning.</td>
<td>$I_{AF} &gt; I_{GK}$ (1)</td>
<td>The resistance of branch AF is less than the resistance of branch GK (0.5), because the HJK network of parallel resistances is in series with another resistance, whereas the BCEF network of parallel resistances is not (0.5). The branch with the least resistance would have the most current flow passing through it (0.5). Hence, the current flow passing through branch AF would be greater than that of branch GK. [1.5]</td>
<td>$I_{AF}$ is equal to $I_{GK}$ (0), because the current flow coming out of the battery when it reaches node $A^*$ splits in half (0). This means that both branch AF and branch GK would have the same current flow passing through them. [0, PPP, 22]</td>
</tr>
</tbody>
</table>

Note. *Students who used PM envisioned the current flow as a continuous loop from the positive terminal of the battery to the negative one, whereas students who used VM knew that the opposite was the case because VM provided them with a view of the current or charge flow.*