Learning and Instruction 21 (2011) 317–331

Learning and Instruction

www.elsevier.com/locate/learninstruc

Physical versus virtual manipulative experimentation in physics learning

Zacharias C. Zacharia*, Georgios Olympiou

Learning in Science Group, Department of Educational Sciences, University of Cyprus, P. O. Box 20537, Nicosia 1678, Cyprus

Received 20 August 2009; revised 18 January 2010; accepted 5 March 2010

Abstract

The aim of this study was to investigate whether physical or virtual manipulative experimentation can differentiate physics learning. There were four experimental conditions, namely Physical Manipulative Experimentation (PME), Virtual Manipulative Experimentation (VME), and two sequential combinations of PME and VME, as well as a control condition (i.e., traditional instruction with absence of PME or VME). Undergraduate students' understanding of physics concepts in the domain of heat and temperature was tested in a pre- and posttest design that involved 182 participants assigned to the four experimental groups and 52 participants assigned to the control group. Conceptual tests were administered to assess students' understanding before, during and after instruction. The analyses revealed that the four experimental conditions were equally effective in promoting students' understanding of concepts in the domain of heat and temperature and better than the control condition; hence, manipulation, either physical or virtual manipulation, and not physicality, as such, at least in a context like the one of the present study, is important in physics learning.

Keywords: Simulations; Virtual labs; Laboratory; Experimentation; Touch; Physics

1. Introduction

Contemporary research focusing on reform in science education is emphasizing the importance of rethinking the role and practice of experimentation in science teaching and learning (Bybee, 2000; National Research Council [NRC], 2000), including physics teaching and learning (Heron & Meltzer, 2005). One of the reasons making this an imperative need is the exponential growth of Virtual Manipulative Experimentation (VME) and its implications for teaching, learning and research in physics. The VME involves the use of virtual apparatus and material that exist in virtual environments, such as computer-based simulations. During the past decade a series of empirical studies revealed the potential of VME to enhance students' skills, attitudes, and understanding of physics concepts (van der Meij & de Jong, 2006; Triona & Klahr, 2003; Zacharia, 2003, 2005; Zacharia & Anderson, 2003). In spite of these findings, some researchers have begun to seriously question whether laboratory experimentation, as we experience it through the use of Physical Manipulative Experimentation (PME), that is, use of real-world physical/concrete material and apparatus, should be redefined and restructured to include VME (Finkelstein et al., 2005; Triona & Klahr, 2003; Zacharia, 2007).

In contrast to the popularity and potential advantages that the use of VME might contribute to physics learning, in general, other researchers, educators and policy makers dispute the use of VME on the grounds that it deprives students of experiences involving physical, "hands-on" manipulation of concrete materials which is essential for learning (Flick, 1993; National Science Teachers Association, 1999). In other words, physical manipulation (actual touch of material and apparatus) is proposed as a prerequisite for physics learning through experimentation. This implies that enacting physics experimentation without the involvement of students in physical manipulation of materials and apparatus, as in the case of VME, could restrict students' learning.

On the other hand, the VME advocates claim, in response to the aforementioned argument, that it is manipulation, rather

* Corresponding author. Tel.: +357 22 753780; fax: +357 22 377950.
E-mail address: zach@ucy.ac.cy (Z.C. Zacharia).

0959-4752/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.learninstruc.2010.03.001

than physicality (i.e., relating to material world with qualities as actually perceived), that may be important for learning through experimentation (Clements, 1999; Klahr, Triona, & Williams, 2007; Resnick, 1998; Triona & Klahr, 2003). Hence, the VME advocates dispute physicality as a requirement for learning. In fact, they argue that experimentation could be enacted without the presence of physicality, through the use of VME, and contribute to students' learning in a similar way as PME, given that VME allows the same affordances in the manipulation of materials and apparatus (Triona & Klahr, 2003), or even in a better way, when VME carries more affordances (e.g., manipulation of reified/conceptual objects such as atoms, vectors etc.) than PME (Zacharia, Olympiou, & Papaevripidou, 2008).

In addition to the PME and VME supporters, there are researchers who advocate in favor of combining the use of PME and VME. In this case, part of an experiment is conducted with PME and the rest with VME. The reasoning behind this mode of experimentation is the reaping of the benefits of both kinds of manipulation (Toth, Klahr, & Chen, 2000; Winn et al., 2006; Zacharia et al., 2008). However, even in this case, physicality, if it proves to be a requirement for learning, is still an issue because when VME is used, the disadvantage (absence of physicality) that it carries is still present, thus, negatively affecting students' learning. It might be the case that for a series of experiments within the same context (e.g., different experiment but same subject domain and skills involved) there is no need for physicality to always be present. Nonetheless, there is no research in this domain so far supporting or questioning such an assumption.

1.1. Physical manipulative experimentation (PME)

The PME is a learning experience that involves a process in which students physically manipulate real-world physical/concrete material and apparatus to observe and understand the natural or material world. The primary characteristic of PME which also differentiates it from other modes of experimentation (e.g., VME) is physicality. Physicality is identified as the actual and active touch of concrete material and apparatus. Active touch is often described as a discovery mode that involves intentional actions through which an object's characteristics (e.g., hardness, texture, weight, inertia, geometry/shape, smoothness, slippage, temperature, etc.) are perceived by someone through tactual/haptic sensation (Loomis & Lederman, 1986).

Besides actual and active touching of materials and apparatus (physicality per se) the enactment of PME also requires manipulation of these materials and apparatus. Manipulation is a process that requires from a learner to intentionally interact with material and apparatus in a skilful manner (in the case of PME it involves certain motor skills, different from those involved in VME). In other words, learners move, arrange, operate, or control by hand the materials and apparatus. Manipulation is usually defined and guided by the design and context of an experiment (e.g., through accompanying teaching material or instructions). No PME could be conducted, unless there is both a purposeful exploration through active touching (physicality) and intentional manipulation of materials and apparatus. For instance, if one of the tasks of an experiment is to bring a sample of water in a beaker at a certain temperature (higher than the one it had before) through the use of a hotplate, it is not enough to just touch the beaker, the hotplate or the thermometer. For such a task, someone needs to position the beaker on the hotplate, insert a thermometer within the water and turn on the hotplate for a certain period of time until it reaches the expected temperature.

1.1.1. Physicality and learning: theoretical and empirical underpinnings

Research studies focusing on physicality and its impact on learning provide evidence that physicality forms the basis for conscious memory and learning (Bara, Gentaz, Pascale, & Sprenger-Charolles, 2004; Klatzky & Lederman, 2002; Loomis & Lederman, 1986), including science learning (Jones, Andre, Superfine, & Taylor, 2003; Jones, Minogue, Tretter, Negishi, & Taylor, 2006; Reiner, 1999). For instance, recent research focusing on the integration of haptics[1] in virtual learning environments (e.g., simulations) have shown promising findings concerning the contribution of physicality to students' science learning (Jones et al., 2003, 2006; Minogue & Jones, 2009), including physics learning (Reiner, 1999).

Jones et al.s' (2003, 2006), in their attempt to explain physicality's contribution to learning, have grounded the predictions of their studies on the working memory theoretical framework (Baddeley, 2003; Baddeley & Hitch, 1974; Millar, 1999), as well as on the cognitive load theory (Sweller, 1994; Sweller, van Merrienboer, & Paas, 1998) which also stresses the role of working memory. In particular, Jones et al's (2003, 2006) argued that tactual experiences may reduce the cognitive load of an individual's working memory during learning and thus support more complex understandings. The latter argument is based on the premise that each modality (visual, auditory, touch) has its own processing channel (Burton & Sinclair, 2000; Millar, 1999). This implies that if multiple channels or modalities are employed, for a given amount of information, the cognitive load is reduced because information is distributed to more than one processing channels (Chan & Black, 2006). However, each one of these three modalities has its own functional specificities (Klatzky & Lederman, 2002).

Such a theoretical framework would explain learning through PME in two ways. First, if the sensory channel of touch can transfer the same information with the one transferred through the visual or the auditory channels, then the cognitive load to the visual and auditory storage systems would be reduced thus leaving capacity for the central executive to support processing that allows more complex understandings. For instance, both visual and tactile modalities (e.g.,

[1] Haptics are devices providing simulated tactile and/or kinesthetic feedback.

size, shape, location) contribute to spatial perception (Bara et al., 2004); hence, the input for spatial perception is divided between the two channels thus resulting to decreased cognitive load for each individual channel. Second, if the sensory channel of touch transfers different (but complementary) information to the one transferred through the visual or auditory channels, learning could still be improved because more information is transferred to the individual's working memory without increasing the cognitive load for each individual channel. For example, touch could provide information that relates to the microstructure (e.g., elasticity, viscosity) of an object whereas vision could provide information that relates to the macrostructure (e.g., color) of it (Klatzky, Lederman, & Reed, 1987) thus facilitating better understanding of a multidimensional object/system (Jones et al., 2003).

Even though such a theoretical framework regarding touch seems plausible, there is no research that has tested how combining the touch modality with visual or auditory modalities affects a learner's cognitive load. Reviews on cognitive load, including the most recent ones (de Jong, 2009), do not include any work on touch with respect to the "modality principle". The evidence on the "modality principle" is coming from research that involves only the visual and auditory modalities (Low & Sweller, 2005). Nonetheless, research on the acquisition of reading skills provides indirect evidence that the "modality principle" may apply when the touch modality is combined with any of the other two modalities, particularly the visual one. Over the past five decades, researchers in this domain, who have been using a "multisensory" technique that combines the visual, auditory and/or touch modalities (Bara et al., 2004; Fernald, 1943), have shown that the mixing of modalities enhances learning more than any learning coming from a single modality. In fact, their findings revealed that the combination of all three sensory modalities provided the optimum result (Bara et al., 2004).

Overall, there is no tested theoretical framework that can explain exactly if and how touch affects one's learning. Given the scarcity of research in this domain and the fact that researchers have not formulated definitive conclusions about the positive effect of physicality, as such, on students' learning, including understanding of science concepts (Minogue & Jones, 2009), it is remarkable that the emphasis on physical manipulation is so prevalent among educators, when it comes to learning through experimentation, even nowadays that alternatives exist. One such alternative is virtual manipulation, which has inherent pragmatic advantages and has accumulated a research record that shows its positive impact on learning (van der Meij & de Jong, 2006; Triona & Klahr, 2003; Zacharia et al., 2008).

One possible explanation for the emphasis on PME is that studies showed in the past PME to be more conducive to students' learning of science concepts than more passive modes of instruction (Glasson, 1989). Such findings were extrapolated to comparisons between PM and VM experimentation, even though there is no empirical data to support such an extrapolation. Moreover, findings provide evidence for the influence of physical manipulation, as a whole, and not of physicality alone (Olympiou, 2006). Therefore, the argument that physicality, as such, is a prerequisite for science learning, is not grounded even in studies comparing PME with more passive modes of instruction.

1.2. Virtual manipulative experimentation (VME)

In the present study, virtual manipulatives refer to virtual, interactive dynamic visualizations of apparatus and material, provided through a computer-based simulation (virtual lab). In this respect, VME refers to a learning experience that involves a process in which students virtually manipulate/interact with these virtual materials and apparatus to observe and understand the natural or material world.

Initially, virtual manipulatives emerged as a need to complement physical manipulatives that inherently carried a number of deficiencies within the context of school science experimentation; this possibly explains the reason why PME advocates still consider VME only as a surrogate for PME and not as a viable method of experimentation in its own right (Kirschner & Huisman, 1998). Despite this restrictive framework, some researchers and developers of learning environments foresaw the potential of VME for learning, such as physics learning, and designed and developed VME environments that move beyond the boundaries set by PME advocates. The attempt was to "match" the experimental affordances provided by PME, and even surpass them by providing more affordances/advantages than PME. As a result, there is a number of VME environments across the science subject domains providing representations that appear to be just as personally meaningful to students as PME and even more manageable, "clean", flexible, and extensible than their physical counterparts (Triona & Klahr, 2003). For instance, the use of VME, like the use of PME, could provide a perceptual grounding for concepts that might otherwise be too abstract to readily comprehend (Winn et al., 2006); promote an active, "hands-on", problem-solving stance that, in turn, often fosters a deep understanding of a phenomenon (NRC, 1999); and provide effective exposure to experimentation and its corresponding skills (van Joolingen & Zacharia, 2009). Alternatively, the use of VME, unlike the use of PME, could provide affordances, such as portability, safety, cost-efficiency, scaffolding, minimization of error, amplification or reduction of temporal and spatial dimensions, manipulation of reified objects, and flexible, rapid, and dynamic data displays (Hsu & Thomas, 2002). A number of studies using VME showed that the use of VME enhanced student learning more than the use of PME (Finkelstein et al., 2005; Zacharia, 2007). Thus, for the purposes of designing experiments comparing PME and VME, as in the present study, controlling for these affordances is extremely important.

Besides VME affordances, there are three other features that vary across VME environments (e.g., virtual labs), namely, richness and transparency of the content of a VME environment, and fidelity. Richness refers to the amount of information and the multiplicity of the relations a learner can extract from the representations of physical and reified/conceptual objects or phenomena. Transparency refers to how "direct" is the perception of the variables, objects (both physical and reified) and relations of the domain which is presented to learners (Swaak, van Joolingen, & de Jong, 1998). Like PME, VME is considered to be a rich learning environment with a relative low transparency (Swaak & de Jong, 1996). Research shows that the richer or the less transparent is a VME environment the more insufficient it becomes to promote learning (Marshall & Young, 2006).

Fidelity is the level of realism that a virtual manipulation presents to the learner (Scheiter, Gerjets, Huk, Imhof, & Kammerer, 2009). In the case of experimentation, fidelity concerns the degree of similarity between VME and PME. Many researchers have studied the relationship between fidelity and learning (Alessi, 1988; Dwyer, 1976; Scheiter et al., 2009). On one hand, it was found that a lowered fidelity enabled learners to acquire knowledge (Alessi, 1988), as it left out the over-stimulating and irrelevant details of physical objects and focused only on the to-be-learned structural features (Dwyer, 1976; Mayer, 2005), which keeps the cognitive load at a minimum (Sweller et al., 1998). On the other hand, proponents of high fidelity in VME argue that the resemblance of VME with the real-world object or phenomenon that it depicts may (a) facilitate its recognition in a natural setting as the perception of the phenomenon or object may directly activate the corresponding mental representations acquired by studying the VME earlier on, or (b) provide perceptual scaffolding for abstract concepts that would be difficult to convey otherwise, and may support inference and reasoning processes grounded in perception (Goldstone & Son, 2005; McNeil, Uttal, Jarvin, & Sternberg, 2009).

1.3. Physical and virtual manipulation

The most fundamental difference between VME and PME is physicality, which for VME advocates, appears not to be a specific requirement for learning, unless the target skill is perceptual-motor (e.g., exploring the operation of an electromagnetic see-saw or making precise incisions during dissections; Triona & Klahr, 2003). From a theoretical perspective, proponents of virtual manipulation argue that the claim that physicality is a requirement for learning is not well-grounded in any of the current major learning theories, namely, the constructivist and the cognitive learning theories. Constructivist theory emphasizes the importance of learners taking an active role in their own learning, but it does not specifically require physical manipulation. Cognitive theory focuses on the need for learners to actively process information and practice the target skill (see for details Triona & Klahr, 2003). From an empirical research viewpoint, there is a surprisingly small number of studies that directly tested the effect of physicality on science learning and involved experimental designs (Triona & Klahr, 2003). The first research findings indicate that physicality, as such, appears not to be a prerequisite for specific learning skills (Triona & Klahr, 2003) or tasks (e.g., discovering the combination of features that yield an optimal design for a car that travels the farthest; Klahr et al., 2007). However, one cannot conclude from these studies whether physicality is a prerequisite for understanding physics concepts.

As in PME, VME involves manipulation of material and apparatus, which is important for learning, but the nature of manipulation is rather virtual than physical. However, dealing with virtual manipulation does not change the essence of manipulation per se (Triona & Klahr, 2003); virtual manipulation is still a process, as in the case of physical manipulation, that entails intentional interactions with material and apparatus in a skilful manner. Students can still move, arrange, operate, or control the "same" material and apparatus, as in physical manipulation. The only difference is that in the case of PME these intentional interactions are performed by hand (e.g., by grabbing and hefting), whereas in the case of VME they are performed by virtual means (by pointing and clicking through a computer's mouse). Hence, due to the absence of physicality (actual and active touching), virtual manipulation also differs from physical manipulation in the type of motor skills that are employed during manipulation. However, VME advocates claim that such a sensory input may not be specifically important for learning (Triona & Klahr, 2003). Returning to the example of heating a sample of water with a hotplate, when using a VME, as in the case of PME, the learner still has the potential to follow the same manipulative processes in bringing the water at the targeted temperature (takes the beaker and places it on the hotplate, inserts a thermometer within the water, and turns on the hotplate for a certain amount of time), as well as, to receive the same feedback that relates to the aim of the learning task, which in this case is to reach a certain temperature (reading from the thermometer). Needless to say, PME and VME do not provide to the learner the same overall feedback. In fact, even if you restrict the affordances of VME in providing additional feedback to the one provided through the thermometer (e.g., feedback that is conceptual in nature, such as, water molecule movement), you cannot remove the physical aspect from PME that provides touch sensory input when manipulating the materials of the experiment (e.g., a sense of how cold or warm the water and the hotplate are at the beginning of the experiment). Nonetheless, for the aims of the experimental task of our example, such a touch sensory input appears not to be important for learning how to heat water and measure its temperature.

1.4. The present study

The aim of the present study was to investigate whether PME, or VME, or even partial exposure to physicality, is important in physics learning at the university level and, particularly, in understanding physics concepts in an attempt to shed light on the aforementioned contrasting views over PME and VME. In all conditions (PME alone, VME alone, PME preceeding VME, VME preceeding PME, and traditional instruction), the teaching and learning affordances of PME and VME and the context of their implementation (i.e., the method of instruction, curriculum/teaching materials, instructors, time-on-task, etc.) was controlled. The aim was to include four distinct experimental conditions with no confounds as well as a control condition in the study. The control condition was used to examine the effect of experimental conditions on students' understanding of heat and temperature (*H&T*) concepts, as compared to a traditional, passive mode of instruction that does not require from students any manipulation of physical or virtual materials and apparatus.

1.4.1. Research questions

More specifically, the five conditions of the study were compared to investigate the following research questions:

(a) Does physicality, as such, influence students' understanding of *H&T* concepts?
(b) Does PME alone or VME alone enhance students' understanding of *H&T* concepts more than traditional instruction does?
(c) Does partial exposure to physicality (i.e., use of PME only at specific parts of the experiment) before or after exposure to virtual manipulation influence students' understanding of *H&T* concepts? Specifically, if physicality is a requirement for understanding physics concepts and students are exposed first to PME and subsequently to VME, is the initial exposure to physicality (PME) sufficient to scaffold students' understanding throughout the subsequent VME experimentation? In the case where students are exposed first to VME and subsequently to PME, is the initial absence of physicality that detrimental to students' understanding to make it impossible to conduct the subsequent experiments/activities and to understand the subsequently introduced physics concepts? This is particularly important in the context of a curriculum, like the one used in the present study, which is constructivist in nature – new knowledge is being constructed on the basis of prior understandings.
(d) Is partial exposure to the two modes of experimentation (PME to VME and vice versa) effective as compared to PME and VME alone, given that the nature of motor skills involved in PME and VME is different? Also, is partial exposure to the two modes of experimentation more effective than traditional instruction?

1.4.2. Hypotheses

Given the aforementioned differing arguments of the PME and VME advocates, as well as, the lack of a coherent theory on physicality and its relation to physics learning, contrasting hypotheses about the effect of physicality were formulated.

With respect to the first research question, the PME advocates argue that physicality is important in physics learning and, therefore the participants in the PME alone condition would have better understanding of *H&T* concepts than those in the VME alone condition (Hypothesis 1a). On the other hand, if physicality is not important in physics learning, then there would be no difference between the effects of PME and VME on the understanding of *H&T* concepts (Hypothesis 1b).

Concerning the second research question, it was hypothesized that the PME alone, but not the VME alone, condition would enhance students' understanding of *H&T* concepts as compared to the control condition (Hypothesis 2a). In contrast, according to VME advocates, both PME and VME would enhance students' understanding of *H&T* concepts as compared to the control condition (Hypothesis 2b).

Concerning the third research question, it was hypothesized that the condition of partial exposure to physicality with PME preceeding VME would promote students' understanding of *H&T* concepts more than the condition of VME preceding PME (Hypothesis 3a). The contrasting hypothesis was that the two sequences would not differ in their effects on students' understanding of *H&T* concepts (Hypothesis 3b).

Finally, as regards the fourth research question, from the PME advocates' viewpoint, the PME alone condition would enhance students' understanding of *H&T* concepts more than the condition of partial exposure to physicality with PME preceding VME; moreover, both of them would have a stronger effect than the condition of partial exposure to physicality with VME preceding PME and the control condition; finally, the condition of partial exposure to physicality with VME preceding PME and the control condition would not differ between them (Hypothesis 4a). In contrast, if physicality is not important in physics learning, then all four experimental conditions would be equally effective and more conducive than the control condition to promoting students' understanding of *H&T* concepts (Hypothesis 4b).

2. Method

2.1. Design

A pre-/posttest experimental design was used as shown in Fig. 1. A VME environment of high fidelity was used; it retained the features and interactions of the subject domain of the study as PME did. Additionally, a conscious effort was made to have the same level of richness and transparency in both the PME and VME environments, and to situate both PME and VME within the same instructional context; that is, the same instructors, method of instruction (inquiry), curriculum/teaching materials (Physics by Inquiry; McDermott & The Physics Education Group, 1996, pp. 163–184) and procedures (as defined by the Physics by Inquiry curriculum; e.g., students work in small groups throughout the course) were used. The selection of the Physics by Inquiry curriculum was based on the fact that it enhances undergraduate students' understanding of physics concepts more than more traditional, passive modes of instruction (Redish & Steinberg, 1999).

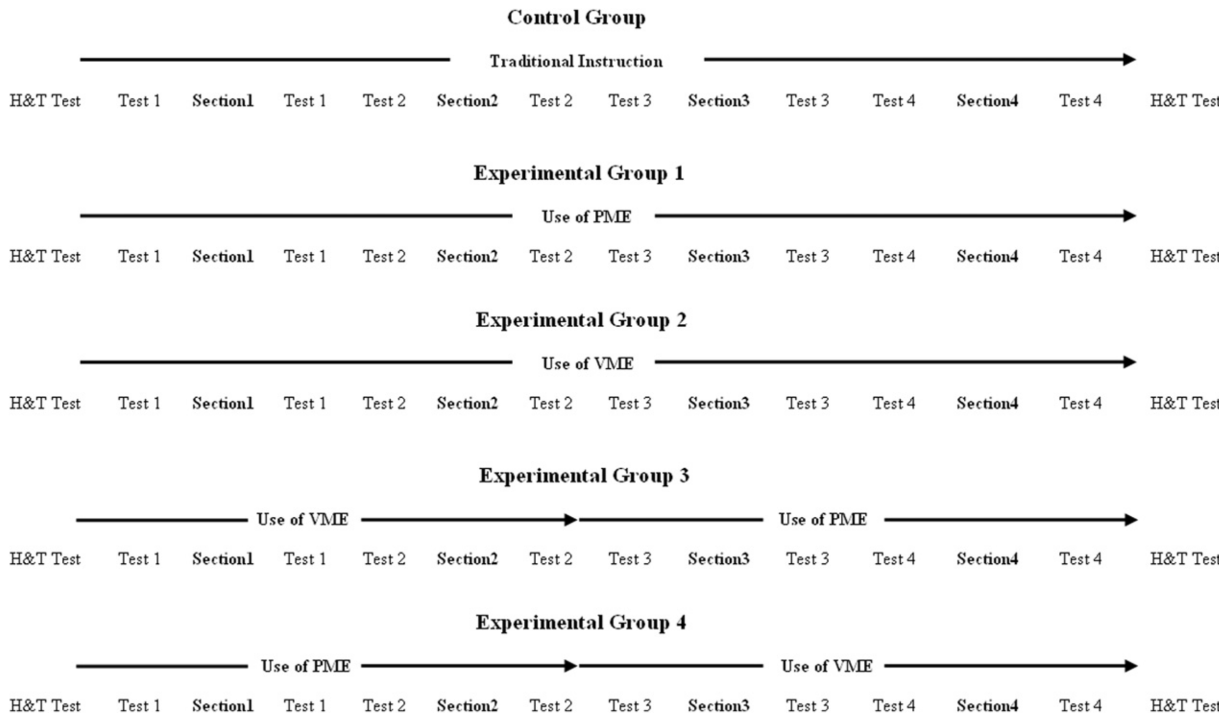

Fig. 1. The experimental design of the study. The sections introduced to the control condition refer to the same topics and concepts as in the experimental conditions.

The experimental study carried out involved the comparison of the effect of PME (i.e., exposure to physicality throughout the study), VME (i.e., no exposure to physicality), two sequential combinations of PME and VME (i.e., partial exposure to physicality) and traditional instruction (control condition, that is, absence of physical and virtual manipulation of materials and apparatus) on undergraduate students' understanding of physics concepts in the domain of *H&T*.

The control condition was derived from the Olympiou (2006) study that aimed at comparing, among others, the effect of PME with that of a traditional mode of instruction within the context of the Physics by Inquiry curriculum on undergraduate students' understanding of concepts concerning temperature, changes in temperature, heat, heat transfer, heat capacity, and specific heat. This former study was contextualized in the same manner as the present study and involved the same data collection processes and sources. In particular, in both the control and experimental conditions the same curriculum/teaching material (same four sections from the Physics by Inquiry curriculum, pp. 163−184) was used. In the case of the control condition, this material was presented to the students through lectures that involved demonstrations of the study's experiments. These demonstrations were made either through the use of videos or real-time demonstrations (both projected on a screen) by the instructor. The experiments included in all demonstrations were conducted through PME. The idea behind the demonstrations was to match, to the highest degree, what the students hear and see in both the control and experimental conditions.

Moreover, the reading material that was available in the control and experimental conditions was matched. Students in the experimental conditions had available as reading material their lab notebook that they were completing themselves as they were progressing with the conduction of the experiments of the curriculum material. What was included in the lab notebook was determined by the curriculum material through specific questions and activities (e.g., construction, interpretation and discussion of a graph). Students in the control condition were given at the beginning of each section a "textbook" (a compilation of notes) that included a description of the experiments to be shown in class and the same accompanying questions and activities as the ones given to the students in the experimental conditions. Students in the control condition were responsible to answer the questions and do the activities themselves during class. They were also allowed and encouraged to discuss their responses to these questions and activities with their peers (as students in the experimental condition did) before having whole class discussions with the instructor. In the case where students needed data to do an activity, such as constructing a graph, ready-made data was provided to them through their "textbook" or by the instructor.

Another variable that was controlled for the comparison of experimental and control conditions was the time-on-task, which was found to be less for the control condition because the projection of a video or the demonstration of an experiment from an expert took less time (sometimes much less time) than the conduction of an experiment from the

Table 1
Time-on-task data for the control group (CG) and the four experimental groups (EG).

Task	Time				
	CG	EG1	EG2	EG3	EG4
	Hours (week)	Hours (week)	Hours (week)	Hours (week)	Hours (week)
H&T test	2 (1)	2(1)	2(1)	2(1)	2(1)
Test 1	1 (2)	1 (2)	1 (2)	1 (2)	1 (2)
Introduction	0.5 (2)	0.5 (2)	0.5 (2)	0.5 (2)	0.5 (2)
Section 1	4.5 (3–5)	4.5 (3–5)	4.5 (3–5)	4.5 (3–5)	4.5 (3–5)
Test 1	1 (5)	1 (5)	1 (5)	1 (5)	1 (5)
Test 2	1 (5)	1 (5)	1 (5)	1 (5)	1 (5)
Section 2	6 (6–9)	6 (6–9)	6 (6–9)	6 (6–9)	6 (6–9)
Test 2	1 (9)	1 (9)	1 (9)	1 (9)	1 (9)
Test 3	1 (9)	1 (9)	1 (9)	1 (9)	1 (9)
Section 3	1.5 (10)	1.5 (10)	1.5 (10)	1.5 (10)	1.5 (10)
Test 3	1 (10)	1 (10)	1 (10)	1 (10)	1 (10)
Test 4	1 (10)	1 (10)	1 (10)	1 (10)	1 (10)
Section 4	6 (11–14)	6 (11–14)	6 (11–14)	6 (11–14)	6 (11–14)
Test 4	1 (14)	1 (14)	1 (14)	1 (14)	1 (14)
H&T test	2 (15)	2 (15)	2 (15)	2 (15)	2 (15)
Total	30.5 h	30.5 h	30.5 h	30.5 h	30.5 h

Experimental Group 1 (EG1): participants used PME alone; Experimental Group 2 (EG2): participants used VME alone; Experimental Group 3 (EG3) participants used both PME and VME with VME preceding the use of PME; Experimental Group 4 (EG4): participants used both PME and VME with PME preceding the use of VME.

students. To compensate for that difference in time (see Table 1 for the study's time-on-task data), in each section of the curriculum material, we used section-related textbook problems (different from the ones used in the tests), which we asked students to solve collaboratively with their peers (they were separated in groups of three). Working collaboratively when solving problems, as well as the fact that the students were encouraged to discuss their responses to the curriculum material's questions and activities with their peers, was an effort to match as much as possible the collaborative practices of the experimental conditions. Overall, the aim of achieving an instructional equivalence between control and experimental conditions was to ensure that in both the control and experimental conditions received the same instructional stimuli.

2.2. Sample

The participants of the study were 234 undergraduate students (66 male, 168 female; $M = 18.5$ years, $SD = 0.9$), enrolled in an introductory physics course at a university in Cyprus, intended for pre-service elementary school teachers. The participants of the control group (CG; 52 students) were derived from a prior study (Olympiou, 2006), whereas, for the rest 182 participants, data were collected in two consecutive semesters. Specifically, 115 participants were randomly separated into two groups, namely, the Experimental Group 1 (EG1, in which participants used PME alone; 56 students) and the Experimental Group 2 (EG2, in which participants used VME alone; 59 students), during the first semester, and 67 participants were randomly separated into two groups, namely, the Experimental Group 3 (EG3, in which participants used both PME and VME with VME preceding the use of PME; 33 students) and the Experimental Group 4 (EG4, in which participants used both PME and VME with PME preceding the use of VME; 34 students), during the second semester (see also Fig. 1). The number of students between the two semesters differed due to the university having offered more places for attending the course during the first semester. None of the participants of all five groups had taken university level physics prior to the study or were attending any other physics course during the study. The tests of the study were completed at a pre-scheduled time outside the course. For this reason, the students who completed all the tests received a bonus toward their final grade in the course.

The Olympiou (2006) study was conducted two years before the conduction of the present study. Nevertheless, in order to reduce the possibility of error because of differences between the participants of the control condition of the former study and those of the experimental conditions of the present study, we compared across all conditions participants' demographics, performance scores across all of the pre-tests (the same pre- and post-tests were used in both studies), as well as the nature and type of their conceptions (the same qualitative analysis was used in both studies). The last two procedures aimed at ensuring that participants in the control group did not differ from those in the experimental conditions in terms of their understanding of concepts concerning temperature, changes in temperature, heat, heat transfer, heat capacity, and specific heat, prior to the study and prior to each section's teaching intervention. As far as the demographics comparison is concerned, it was found that the participants in the control condition and the participants in all experimental conditions (a) shared about the same proportion of men and women, (b) had about the same mean age (and SD), (c) had the same ethnicity, (d) shared about the same K-12 educational background (e.g., all participants in all conditions were graduates of public schools, all groups shared about the same

Table 2
Mean scores (and SD) of the control group (CG) and four experimental groups (EG) in each of the tests.

Group	*H&T* test	Test 1	Test 2	Test 3	Test 4
CG pre-test	34.1 (15.3)	19.8 (8.4)	33.1 (8.8)	24.0 (7.8)	30.8 (13.7)
CG post-test	44.6 (12.7)	52.0 (12.3)	54.8 (10.2)	36.2 (8.7)	55.1 (13.0)
EG1 pre-test	34.6 (11.7)	21.5 (8.9)	34.5 (7.6)	23.7 (8.4)	31.9 (11.0)
EG1 post-test	68.4 (14.0)	75.3 (12.6)	70.3 (10.2)	56.0 (13.9)	80.4 (19.3)
EG2 pre-test	33.1 (10.5)	18.0 (9.0)	32.2 (8.7)	25.1 (8.0)	34.0 (12.8)
EG2 post-test	68.9 (13.2)	74.8 (16.7)	69.0 (8.9)	57.2 (14.4)	79.7 (18.4)
EG3 pre-test	35.0 (15.1)	20.8 (8.8)	35.4 (7.5)	22.8 (7.8)	31.0 (12.3)
EG3 post-test	66.4 (16.4)	79.8 (15.7)	71.1 (12.4)	58.0 (12.2)	77.0 (17.4)
EG4 pre-test	34.8 (12.1)	21.4 (6.3)	35.0 (6.3)	24.5 (9.2)	29.6 (10.3)
EG4 post-test	65.3 (14.6)	75.1 (14.4)	70.7 (11.0)	58.7 (12.6)	80.0 (14.8)

Experimental Group 1 (EG1): participants used PME alone; Experimental Group 2 (EG2): participants used VME alone; Experimental Group 3 (EG3) participants used both PME and VME with VME preceding the use of PME; Experimental Group 4 (EG4): participants used both PME and VME with PME preceding the use of VME.

distribution of participant scores on the university entrance exams), and (e) were at the same (second) year of their undergraduate studies. Finally, the quantitative and qualitative analyses showed that the participants in both experimental and control conditions shared similar understanding of *H&T* concepts. The quantitative analysis (one-way ANOVA) showed that the participants' performance scores in the control condition, on the same pre-tests that were administered in both studies, were not significantly different from the scores of the participants in the experimental conditions, $F < 1$, ns (see Table 2 for means and standard deviations), and the qualitative analysis showed that the nature and type (scientifically accepted or not) of students' conceptions did not differ, across all of the categories of concepts investigated: for temperature, $\chi^2(4, N = 234) = 0.33, p > 0.05$; for changes in temperature, $\chi^2(4, N = 234) = 2.01, p > 0.05$; for heat, $\chi^2(4, N = 234) = 0.4, p > 0.05$; for heat transfer, $\chi^2(4, N = 234) = 1.31, p > 0.05$; for heat capacity, $\chi^2(4, N = 234) = 1.53, p > 0.05$; and for specific heat, $\chi^2(4, N = 234) = 5.6, p > 0.05$.

2.3. Curriculum materials

For the aims of this study the first four sections of the module of Heat and Temperature (*H&T*) of the Physics by Inquiry curriculum were used.[2] The first section (Section 1) focuses on constructing an operational definition for temperature, the second section (Section 2) focuses on investigating temperature changes when samples of hot and cold water are mixed, the third section (Section 3) introduces concepts concerning heat and heat transfer in the context of two objects of different temperatures that interact thermally. Moreover, the third section focuses on clarifying/distinguishing the role of heat and temperature in thermal interactions. The fourth section (Section 4) focuses on the thermal properties of matter, particularly, heat capacity and specific heat capacity. In these four sections, the students are encouraged (a) to make the necessary mental commitment by guiding them through the process of constructing a conceptual model for how temperature changes starting from direct "hands-on" experience that involves mixing varying amounts of hot water and cold water, as well as, water and other substances, and (b) to develop the concepts necessary to describe matter in terms of its thermal properties.

2.3.1. Physical manipulative experimentation (PME)

The PME involved the use of real instruments (thermometers), objects (heaters, beakers, and Styrofoam cups), and materials (wood, aluminium, and water) for constructing an experimental set-up in a conventional physics laboratory. During PME feedback is available to the students through the behaviour of the actual system (e.g., water boiling or not) and through the instruments that are used to monitor the experimental set-up (e.g., thermometers).

2.3.2. Virtual manipulative experimentation (VME)

The VME involved the use of virtual instruments (thermometers), objects (heaters, beakers, and Styrofoam cups), and materials (wood, aluminium, and water) to conduct the study's experiments on a computer. In this study, the Virtual Lab ThermoLab (see Fig. 2) was used for this purpose (Hatzikraniotis et al., 2001). ThermoLab was selected because of its fidelity and the fact that it retained the features and interactions of the domain of *H&T* as PME did. In its open-ended environment, students using VME were able to design and conduct the same experiment mentioned in the module of *H&T* by employing the "same" material and equipment as the ones used by the students in the PME condition.

Students were able to construct their own virtual experimental set-ups by simple and direct manipulation of objects, materials and virtual instruments. The software offered feedback throughout the conduct of the experiment by presenting information (e.g., time, temperature, volume) through the displays of the software and by animating the phenomenon (e.g., bubbles and steam come out of solution when water is boiled). No feedback was provided by the software during the

[2] The wording of the curriculum/teaching material used in the VME condition was slightly modified to refer to the features of the virtual manipulatives.

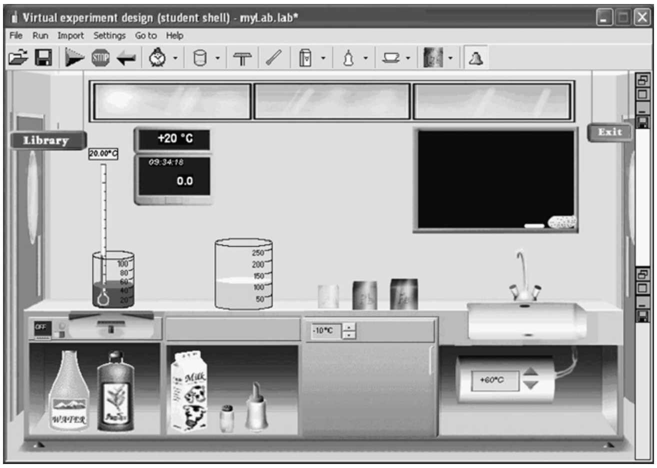

Fig. 2. ThermoLab.

preparation of the (virtual) experimental set-up. The level of feedback was analogous to what is routinely available to students through PME.

2.4. Tasks – measures

The same conceptual test (*H&T* test) was administered to assess students' understanding of *H&T* concepts concerning temperature, changes in temperature, heat, heat transfer, heat capacity, and specific heat both before and after the study. Additionally, tests specific to each section of the study were administered before and after introducing each section (Test 1, 2, 3 and 4; see Fig. 1), with each test being identical before and after each section. The tests were developed and used in previous research studies by the Physics Education Group of the University of Washington (McDermott & The Physics Education Group, 1996), as well as, by our own research group (Olympiou, 2006; Zacharia & Constantinou, 2008; Zacharia et al., 2008). Each of the Tests 1, 2, 3 and 4 contained four items (some of which consisted of two sub-items and each one of these sub-items of at least one question; see Appendix A for a sample of an item of Test 2) that asked open-ended conceptual questions all of which required explanations of reasoning. The *H&T* test included eight open-ended items assessing all sections of the study's curriculum. This test targeted both the specific concepts introduced in each section, as well as, the interconnections and interdependencies of these concepts. No identical items were included in the *H&T* test and the rest of the tests.

Each item of each test was scored separately; however, for correct responses a total score was derived from each test and used in the analysis. All tests were scored and coded blind to the condition in which the student was placed. The scoring of each item was performed through the use of a scoring rubric table that included pre-set criteria (expected correct answer and expected correct explanation; see Appendix A for an example), which were used to score both whether a participant's answer to a question of an item and its accompanied explanation were correct. A correct answer to a question was always scored with one point, across all tests, and its corresponding explanation in accordance with how many of its preset criteria (expected knowledge needed to explain an answer) were met. Each explanation criterion was scored with half point, across all tests. The maximum score of each question of an item of a test varied according to the number of criteria used for scoring its explanation. Hence, the maximum score of an item of a test varied both across the items of a test and across the items of the other tests, unless two items shared the same total number of explanation criteria. An individual's total score on a test was derived by adding all the assigned scores, both those of an answer and an explanation, of all questions (of all items) of a test, and by adjusting it to a 100-point scale. The minimum score was 0 and the total maximum score was 100 on each test. Two independent raters scored about 20% of the data. The reliability measures (Cohen's kappa) for scoring of the *H&T* test (pre- and posttest) and Tests 1, 2, 3 and 4 (pre- and posttests), were 0.89, 0.90, 0.92, 0.88, and 0.90, respectively. The reliability (proportion of agreement) of the scoring of the qualitative data (students' conceptions; e.g., "Suppose equal volumes of iron and water interact thermally. Which will change temperature more? Explain your reasoning") was 0.91. Disagreements were discussed after the reliability analysis, and were classified when mutual agreement was reached.

2.5. Procedure

Despite the fact that the data collection occurred at different times, the procedures followed were identical at all times. First, all groups were formed after random assignment of the participants to a particular condition. The CG students attended the course in one of the university's lecture classes, whereas all EG students worked in the same laboratory environment that hosts both conventional equipment and a computer network arranged at the periphery.

Second, within each condition students were randomly assigned to subgroups (of three) as suggested by the curriculum of the study (in some groups there were one or two 2-member subgroups because the total number of participants in the group was not enough to form triads). In the case of the control group, students worked in subgroups of three only during the solution of textbook problems.

Third, all participants were administered the *H&T* pre-test before getting engaged in the treatment of the condition they belonged to. After a week, Test 1 was administered and a brief introduction that aimed to familiarize students with the material they were about to use. The CG students were introduced to the "textbook" that was prepared for them and the procedures that they were going to follow throughout the course. All EG students were introduced to the Physics by Inquiry curriculum and to both VME and PME through a demonstration, regardless of their condition. The introduction to the routines and procedures of the Physics by Inquiry curriculum was very important because they differ from those involved in the more traditional, passive modes of instruction that students had experienced in physics courses during their

K-12 years. For example, the enactment of the Physics by Inquiry curriculum does not involve any lecturing, tutoring, or traditional textbook. In contrast, students are seen as responsible for their own learning and are expected to collaboratively construct knowledge and develop their understanding of physics concepts through the conduction of a carefully designed, structured sequence of inquiry-based experiments.

Moreover, the role of the instructors in the Physics by Inquiry curriculum is quite different from that in a traditional instruction. It is supportive in nature and requires instructors' engagement in dialogues with the students of a subgroup at particular points of the activity sequence, as specified by the Physics by Inquiry curriculum. Through these dialogues, the instructors aim to encourage reflection across the inquiry processes and practices involved in the activities of the Physics by Inquiry curriculum and not to lecture or provide ready-made answers/solutions. For consistency in the instructors–students interactions and dialogues, all instructors have meetings on a weekly basis that focus on reviewing the material/activities before these are encountered in class and possible issues or prompts that should be discussed or used, respectively, during the dialogues with the students. For the purposes of this experiment, EG1 and EG2 students shared the same seven instructors (consisted of one academic and six doctoral students) throughout the instructional intervention, whereas, EG3 and EG4 students shared the same five instructors (same academic as in the other two EG and four out of the six doctoral students that were used in the other two EG). All instructors were previously trained in implementing the Physics by Inquiry curriculum and had experienced its implementation at least for two years. The instructor, in the case of CG, was the same academic who was involved in the EG conditions.

Fourth, along with the instructional part of each section, conceptual tests were also administered both before and after each section (see Fig. 2).

The duration of the study was 15 weeks. Students met once a week for one and a half hour. The time-on-task was the same for all groups (see Table 1 for the time-on-task data). In doing so, we controlled for any deviations between the time-on-task across all conditions (it was found elsewhere to affect student learning; see for an example Zacharia et al., 2008), particularly among the experimental ones due to a difference in the possibilities afforded by PME and VME for heating materials. For example, because it takes students using physical manipulatives much more time to bring a sample of water to a certain temperature than the students using VME, the groups that used PME were provided with pre-set material (e.g., pre-heated samples of water) to avoid any time consumption on such routine procedural tasks. To ensure that the time-on-task variable could be controlled successfully, a pilot study was run right before the present study (see Zacharia & Constantinou, 2008).

2.6. Data analysis

All tests were scored though the use of scoring rubrics and the resulted student performance scores were analyzed by using (a) one-way ANOVA for the comparison of the pre-test scores of the five groups on each test and (b) one-way ANCOVA for the comparison of the post-test scores of the five groups on each test. For the latter procedure, the students' scores in the corresponding pre-tests were used as the covariate. The aim of the first procedure was to determine whether the five groups of the study were comparable with regard to the sample's entry understanding of physics concepts from the subject domain of *H&T*, before the study and before each section. The aim of the second procedure was to investigate whether the five groups of the study had differences on the outcome measures (understanding of physics concepts in the domain of *H&T*) of each test. For all analyses the effect size (partial η^2; Cohen, 1988) is reported.

The second way involved the identification and classification of students' scientific and non-scientific conceptions concerning temperature, changes in temperature, heat, heat transfer, heat capacity, and specific heat through the use of qualitative thematic analysis. A sample of the qualitative analysis concerning conceptions of specific heat is included in Table 3.

In addition, the percentage for each one of the resulting conceptions was calculated. The aim of the latter was to compare whether the prevalence of students' conceptions changed after the treatments of the study. This procedure was essential because it clarified whether students of similar scores shared the same ideas or not, either scientifically acceptable or scientifically not acceptable ones.

3. Results

3.1. Test performance

Means and standard deviations of performance scores are shown in Table 2.

Students' scores to the *H&T* post-test were subjected to an ANCOVA with *H&T* pre-test scores as covariate and group as between subjects factor. The analysis revealed a main effect of group, $F(4, 228) = 52.3$, $p < 0.001$, partial $\eta^2 = 0.48$, and of *H&T* pre-test scores, $F(1, 228) = 116.8$, $p < 0.001$, partial $\eta^2 = 0.36$, but no interaction between group and *H&T* pre-test scores, $F < 1$, ns.

In the case of Test 1, the ANCOVA revealed a main effect of group, $F(4, 228) = 31.5$, $p < 0.001$, partial $\eta^2 = 0.36$, and of pre-test 1 scores (covariate) on students' post-test 1 scores, $F(1, 228) = 29$, $p < 0.001$, partial $\eta^2 = 0.11$, but no interaction between group and pre-test scores, $F < 1$, ns.

In the case of Test 2, the ANCOVA revealed a main effect of group, $F(4, 228) = 23.3$, $p < 0.001$, partial $\eta^2 = 0.30$, and of pre-test 2 scores (covariate) on students' post-test 2 scores, $F(1, 228) = 36.7$, $p < 0.001$, partial $\eta^2 = 0.13$, but no interaction between group and pre-test scores, $F < 1$, ns.

In the case of Test 3, the ANCOVA also revealed a main effect of group, $F(4, 228) = 52.3$, $p < 0.001$, partial $\eta^2 = 0.48$, and of pre-test 3 scores (covariate) on students' post-test 3 scores, $F(1, 228) = 119.5$, $p < 0.001$, partial $\eta^2 = 0.36$, but no interaction between group and pre-test scores, $F < 1$, ns.

Table 3
A sample of students' conceptions about specific heat capacity as they emerged from the qualitative thematic analysis of Test 4.

Conceptions	CG ($n = 52$)		EG1 ($n = 56$)		EG2 ($n = 59$)		EG3 ($n = 33$)		EG4 ($n = 34$)	
	Pre-tests	Post-tests	Pre-tests	Post-tests	Pre-tests	Post-tests	Pre-tests	Post-tests	Pre-tests	Post-tests
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
1. The amount of heat that changes the temperature of 1 g of an object by 1°C is always the same, or the Specific Heat of an object indicates the amount of heat that will change the temperature of 1 g of the object by 1°C (SAC)	0 (0)	26.9 (14)	3.5 (2)	89.2 (50)	6.7 (4)	94.9 (56)	3 (1)	84.8 (28)	0 (0)	91.1 (31)
2. The amount of heat that changes the temperature of objects, of the same substance, by 1°C is always the same (SNAC)	82.6 (43)	61.5 (32)	62.5 (35)	9.0 (5)	79.6 (47)	5.1 (3)	81.8 (27)	15.1 (5)	70.5 (24)	8.9 (3)
3. The amount of heat that changes the temperature of 1 g of an object by 1°C is the same to the amount of heat that changes the temperature of the entire object by 1°C (SNAC)	7.7 (4)	3.8 (2)	19.6 (11)	1.7 (1)	11.8 (7)	0 (0)	9.0 (3)	0 (0)	14.7 (5)	0 (0)
4. The amount of heat that changes the temperature of 1 g of an object by 1°C depends upon the initial temperature of the object (SNAC)	9.6 (5)	7.7 (4)	12.5 (7)	0 (0)	1.7 (1)	0 (0)	3.0 (1)	0 (0)	5.8 (2)	0 (0)
5. The amount of heat that changes the temperature of 1 g of an object by 1°C varies (SNAC)	0 (0)	0 (0)	1.7 (1)	0 (0)	0 (0)	0 (0)	3.0 (1)	0 (0)	8.8 (3)	0 (0)

SAC = Scientifically acceptable conception. SNAC = Scientifically not acceptable conception. The percentages and number (in parenthesis) 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. In addition, the response of one individual student may appear in more than one conception depending on the conceptions that were evident in the response. Experimental Group 1 (EG1): participants used PME alone; Experimental Group 2 (EG2): participants used VME alone; Experimental Group 3 (EG3) participants used both PME and VME with VME preceding the use of PME; Experimental Group 4 (EG4): participants used both PME and VME with PME preceding the use of VME.

Finally, in the case of Test 4, the ANCOVA revealed a main effect of group, $F(4, 228) = 26.7$, $p < 0.001$, partial $\eta^2 = 0.32$, and of pre-test 4 scores (covariate) on students' post-test 4 scores, $F(1, 228) = 80.4$, $p < 0.001$, partial $\eta^2 = 0.24$, but no interaction between group and pre-test 4 scores, $F < 1$, ns.

Bonferroni-adjusted ($p < 0.01$) pairwise comparisons suggested that students' post-test scores in the control group were significantly lower than those of students in the four experimental conditions across all tests. The pairwise comparisons did not show any significant difference between the students' post-test scores of the experimental groups across all tests. These findings suggest that the use of PME alone, VME alone, and the two sequential combinations (VME preceding PME and PME preceding VME) enhanced students' understanding of the *H&T* concepts more than traditional instruction did; moreover, that all the experimental conditions were equally effective in promoting students' understanding of these concepts.

3.2. Understanding of H&T concepts

The qualitative analysis revealed that all of the experimental groups shared mostly the same conceptions across the *H&T* concepts studied (i.e., temperature, changes in temperature, heat, heat transfer, heat capacity, and specific heat), as either scientifically acceptable (SAC) or scientifically not acceptable (SNAC) conceptions, both before and after the *H&T* test was administered. The same regarded Tests 1, 2, 3 and 4 both before and after the introduction of each section of the study's curriculum. The control group appeared to share the same SAC and SNAC conceptions with the experimental groups only before the study at the pre-test of each section (see for an example Table 3).

Overall, in the *H&T* conceptual test most of the students of the experimental groups shifted from SNAC to SAC ($M = 147.3$, $SD = 5.2$) across the *H&T* concepts investigated, after the completion of the four sections. The experimental groups were found to have higher prevalence for each SAC and lower prevalence for each SNAC across all post-tests than the control group (see Table 4). The prevalence of each SNAC and SAC of the participants of the control group was found to be about the same at the *H&T* pre- and posttest, as well as at each test before and after each of the four sections. Lastly, all groups were found to share the same most prevalent SNAC across all pre- and posttests. Finally, these findings indicate that the use of PME and VME, alone or in sequential combination, had the same effect on undergraduate students' understanding of *H&T* concepts, namely, on the transition from SNAC to SAC, as well as, on the type of conceptions students had after the completion of each section. For the mean frequencies and standard deviations of SAC and SNAC conceptions in the five groups and in the five tests at the pre- and posttest see Table 4.

4. Discussion

The aim of the present study was to investigate whether physicality or manipulation (physical or virtual) is important in physics learning at the university level, and particularly in understanding of physics concepts. The findings of this study

Table 4
The mean frequencies and standard deviations of SAC and SNAC conceptions in the five groups and in the five tests at the pre- and posttest.

Test	Conception Type	CG (n = 52) Pre-tests M (SD)	CG (n = 52) Post-tests M (SD)	EG1 (n = 56) Pre-tests M (SD)	EG1 (n = 56) Post-tests M (SD)	EG2 (n = 59) Pre-tests M (SD)	EG2 (n = 59) Post-tests M (SD)	EG3 (n = 33) Pre-tests M (SD)	EG3 (n = 33) Post-tests M (SD)	EG4 (n = 34) Pre-tests M (SD)	EG4 (n = 34) Post-tests M (SD)
H&T test	SAC	4.8 (2.1)	9.1 (3.1)	4.9 (1.6)	15.1 (3.5)	4.7 (1.5)	15.3 (3.3)	5.0 (2.1)	14.8 (4.1)	4.8 (1.7)	14.7 (3.4)
	SNAC	11.9 (3.8)	8.6 (4.6)	11.8 (4.1)	5.7 (3.4)	11.6 (3.3)	6.0 (3.0)	12.0 (5.1)	6.1 (3.6)	11.4 (4.0)	6.2 (3.1)
Test 1	SAC	0.9 (0.3)	2.1 (0.6)	0.9 (0.4)	3.4 (0.5)	0.8 (0.4)	3.3 (0.7)	1.0 (0.4)	3.6 (0.7)	0.9 (0.5)	3.3 (0.6)
	SNAC	4.3 (1.9)	3.2 (1.0)	4.1 (2.1)	1.0 (0.9)	3.9 (1.8)	1.4 (1.2)	4.7 (2.4)	1.3 (1.7)	3.9 (2.0)	1.6 (1.9)
Test 2	SAC	1.5 (0,4)	2.4 (0.5)	1.6 (0,3)	3.4 (0.4)	1.5 (0.6)	3.4 (0.4)	1.4 (0.6)	3.2 (0.6)	1.6 (0.2)	3.3 (0.4)
	SNAC	4.0 (1.8)	3.2 (1.4)	3.9 (1.8)	2.5 (1.4)	3.7 (1.8)	2.6 (1.3)	3.8 (2.0)	2.8 (1.8)	3.5 (2.3)	2.7 (1.2)
Test 3	SAC	1.0 (0.3)	1.6 (0.4)	1.1 (0.4)	2.7 (0.6)	1.2 (0.3)	2.9 (0.7)	1.0 (0.3)	2.6 (0.5)	1.1 (0.4)	2.8 (0.6)
	SNAC	2.5 (0.9)	1.9 (1.1)	2.6 (1.0)	1.3 (1.4)	2.8 (1.0)	1.0 (1.1)	2.6 (0.8)	0.9 (0.8)	2.4 (1.1)	1.0 (0.9)
Test 4	SAC	1.4 (0.6)	2.4 (0.5)	1.6 (0.5)	3.7 (0.9)	1.5 (0.6)	3.6 (0.8)	1.4 (0.6)	3.5 (0.8)	1.3 (0.4)	3.6 (0.7)
	SNAC	3.2 (1.0)	2.2 (1.7)	3.1 (1.4)	0.9 (1.1)	3.2 (1.2)	1.1 (1.0)	3.3 (1.7)	1.3 (1.3)	3.4 (1.5)	1.2 (1.0)

SAC = Scientifically acceptable conception. SNAC = Scientifically not acceptable conception. Experimental Group 1 (EG1): participants used PME alone; Experimental Group 2 (EG2): participants used VME alone; Experimental Group 3 (EG3) participants used both PME and VME with VME preceding the use of PME; Experimental Group 4 (EG4): participants used both PME and VME with PME preceding the use of VME.

indicate that the use of PME and VME, either used alone or in sequential combination, when embedded in a context similar to the one of this study, can equally enhance students' understanding of *H&T* concepts and more than traditional instruction. These findings confirm Hypotheses 1b, 2b, 3b and 4b, which are in line with previous findings (Hofstein & Lunetta, 2004).

These findings also challenge the general assumption of the PME advocates that physicality is a requirement for physics learning. None of the predictions (Hypotheses 1a, 2a, 3a, 4a) based on this assumption was verified. In contrast, the findings suggest that what is important in physics learning is manipulation, physical or virtual, rather than physicality, at least within a context similar to the one of the present study. Of course, this finding does not provide conclusive evidence that physicality, in general, is not a requirement for individuals' understanding of physics concepts, or that the working memory model described above (Millar, 1999) and its effect on cognitive capacity and load is not valid. To reach such conclusions further research is needed that covers physics learning across K-12.

Moreover, the design of future studies should allow testing of hypotheses about the sensory channels actively used during experimentation, and how this use of sensory channels affected students' cognitive load as well as conceptual integration of multimodal information. Factors, such as the age of the participants or, even better, prior exposure to PME through everyday life and K-12 experiences, also need to be investigated. It might be the case, for example, that the participants of the present study who used VME did not need the sensory input from touch because this information was already in students' long-term memory from prior everyday life or K-12 experiences. Even if the latter assumption were true, the finding of the present study that manipulation appears to be the important aspect of understanding physics concepts at the university level, and not physicality as such, is still valid. Nonetheless, we still need to know if prior exposure to physicality is necessary or not for understanding certain physics concepts. In case it proves to be necessary, it is important to know when the students need to be exposed to this touch-related information. Furthermore, the implications of such a finding for the curriculum material and learning environments need to be delimited. Therefore, further research is needed to illuminate the mechanism through which physicality affects physics learning.

A second aim of the present study was to investigate whether exposure to partial physicality, that is, whether combining the mode of manipulation (physical to virtual or vice versa) in the same sequence of instructional activities as in PME alone and in VME alone would have a differential effect on students' conceptual understanding compared to PME alone and VME alone; also, to examine whether the effect is different when physical manipulation precedes virtual manipulation and vice versa. The findings indicated that the two sequential combinations, in which the mode of manipulation was switched, did not differ between them and from PME alone and VME alone, thus verifying Hypothesis 4b. The fact that the switching of mode of experimentation could occur without affecting students' understanding provides support to Triona and Klahr's (2003) claim that the sensory input coming from the corresponding manipulation or motor skills may not be specifically important for learning. What appears to be important is whether the essential variables and interactions are retained the "same" between physical and virtual manipulation conditions. Again, this conclusion needs to be further investigated, particularly if someone considers that the motor skills employed in both modes of manipulation were very simple and, most probably, have already been used by the students during K-12. For instance, there is need to investigate questions such as "Is the switching of the manipulation mode feasible when the motor-skills involved in the physical mode are complex (e.g., dissections)?", or "How prior experience with the physical or virtual manipulation motor skills influence the effectiveness of switching of the manipulation mode?"

4.1. Educational implications

The findings of the present study also provide information about the potential and value of the use of PME and VME in physics learning, particularly of VME which has been disputed as a viable means for learning as opposed to PME. Both the quantitative and qualitative analysis of the present study's data showed that the use of VME promoted students' understanding of physics concepts equally well as PME, with the provision that PME and VME are implemented within a context similar to the one of the present study. This finding supports what has been argued in recent studies about the comparable effectiveness of PME and VME in promoting science learning (Klahr et al., 2007; Triona & Klahr, 2003; Zacharia & Constantinou, 2008).

An interesting finding coming out of the qualitative analysis of the present study's data that further supports the above argument is that the vast majority of the students in all experimental conditions appeared to share the same SNAC conceptions, both in the pre- and posttests. This finding demonstrates that the nature of learning and the learning outcomes do not substantially change when PME is substituted with VME. This finding provides further credence to the idea that VME can be used (in some contexts and given specific conditions) to provide authentic laboratory experiences that are not substantially different to the methods employed when using PME.

The question raised at this point is which of the two modes of experimentation should be preferred when PME and VME offer the same affordances for physics learning through experimentation, as in the present study. Obviously, any of the VME and PME learning environments would do. Nonetheless, if a teacher has to choose between VME and PME other factors besides the affordances of each of the two types of learning environments should be considered. For example, issues of portability, safety, or cost-efficiency could be considered. Klahr et al. (2007) pointed to several of these "external" factors that a teacher could consider. For example, they argued that VME usually takes less space and effort and affords easier classroom management than PME. Also, they presented the easy duplication and distribution of VME as another obvious advantage over physical "science kits" (assuming, of course, that computers are available to run the programs) and, finally, VME are easier than PME to replicate and distribute.

4.2. Limitations of the study

It should be noted that the present study was carried out in the context of a normal course in physics, thus offering ecological validity to the aforementioned findings. On the other hand, the study had some limitations related to the timeframework that was followed for the data collection. First, the data for the control group were derived from a former study of the authors that was conducted about two years before the collection of the data that concerned the experimental groups of the present study and, second, the experimental groups were not all run concurrently. Yet, the students of the control group were not found to differ, either in the nature and type of their conceptions or statistically in their pre-test performance from that of the students in the experimental groups (both studies used the same pre-tests). Additionally, all students followed the same curriculum, Physics by Inquiry, and all students had the same age and educational background. Another limitation of this study is that it involved only one data source (conceptual tests). We could have had a better insight in terms of student learning that took place during our intervention, if we had used more data sources, particularly data sources focusing on the process rather than just the end results. For instance, in the case of VME, we could have taken on-line measures of learning.

However, despite these limitations, the findings of the present study still provide some valuable information. In particular, findings like the ones of this study challenge the already established norms concerning experimentation in the physics classroom, as we experienced it through PME, in a way that calls for a redefinition and restructuring of experimentation to include VME. However, this call for reform creates the need for understanding how PME and VME could be integrated in teaching and learning activity sequences for physics. Therefore, it is essential to extend the empirical base through similar research in order to ground the perspectives advocated in this study.

Appendix A. Item 1 of Test 2

A. Suppose that 500 g of hot water at 60 °C is mixed with 500 g of cold water at 40 °C.
   1. The temperature of the hot water increases, decreases, or remains the same? Explain your reasoning.
   2. The temperature of the cold water increases, decreases, or remains the same? Explain your reasoning.
   3. The temperature of the mixture will be higher than 50 °C, lower than 50 °C, or exactly at 50 °C? Explain your reasoning.

B. Suppose that 250 g of hot water at 60 °C is mixed with 500 g of cold water at 40 °C.
   1. The temperature of the hot water increases, decreases, or remains the same? Explain your reasoning.
   2. The temperature of the cold water increases, decreases, or remains the same? Explain your reasoning.
   3. The temperature of the mixture will be higher than 50 °C, lower than 50 °C, or exactly at 50 °C? Explain your reasoning.

Scoring rubric for Item 1 of Test 2 (score in parenthesis).

Question	Correct answer	Expected explanation	Total score
A1	The temperature of the hot water decreases. (1)	1. The two samples of water are in contact. (0.5) 2. The two samples of water interact thermally. (0.5) 3. The two samples of water have different temperatures. (0.5) 4. Given that the mass of the hot water times the temperature change of the hot water equals the mass of the cold water times the temperature change of the cold water, the temperature of the hot water decreases. (0.5)	3
A2	The temperature of the cold water increases. (1)	1. The two samples of water are in contact. (0.5) 2. The two samples of water interact thermally. (0.5) 3. The two samples of water have different temperatures. (0.5) 4. Given that the mass of the hot water times the temperature change of the hot water equals the mass of the cold water times the temperature change of the cold water, the temperature of the cold water increases. (0.5)	3
A3	The temperature of the mixture will be 50 °C. (1)	1. The mass of the hot water times the temperature change of the hot water equals the mass of the cold water times the temperature change of the cold water, or $M_h.\Delta\theta_h = M_c.\Delta\theta_c$. (0.5) 2. The two samples of water have the same mass, or $M_h = M_c$. (0.5) 3. The two samples of water have different temperatures, or $\Delta\theta_h \neq \Delta\theta_c$. (0.5) 4. The final temperature of the mixture should be the average temperature (50 °C) of the initial temperatures of the two samples of water (40° and 60 °C), or $\theta_{equilibrium} = (\theta_{initial\text{-}c} + \theta_{initial\text{-}h})/2 = (40\ °C + 60\ °C)/2 = 50\ °C$. (0.5)	3

Students were not required to refer to the concept of heat because the concepts of heat and heat transfer (Section 3 of the Heat and Temperature module of the Physics by Inquiry curriculum) are introduced right after the section on Changes in Temperature (Section 2 of the Heat and Temperature module of the Physics by Inquiry curriculum).

The equation $M_h.\Delta\theta_h = M_c.\Delta\theta_c$ is discovered by the students through an experiment of Section 2 that involves mixing various amounts of hot and cold water, and requires plotting a graph of the mass times the temperature change for the hot water versus the mass times the temperature change for the cold water and computing the slope of the graph.

References

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*, 40–47.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic.

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

Burton, H., & Sinclair, R. J. (2000). Attending to and remembering tactile stimuli: a review of brain imaging data and single-neuron responses. *Journal of Clinical Neurophysiology, 17*, 575–591.

Bybee, R. (2000). Teaching science as inquiry. In J. Minstrel, & E. H. Van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 20–46). Washington, DC: American Association for the Advancement of Science (AAAS).

Chan, M. S., & Black J. B. (2006, June). Direct-manipulation animation: Incorporating the haptic channel in the learning process to support middle school students in science learning and mental model acquisition. Paper presented at the International Conference of the Learning Sciences, Bloomington, Indiana, USA.

Clements, D. (1999). "Concrete" manipulatives, concrete ideas. *Contemporary Issues in Early Childhood, 1*, 45–60.

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

Dwyer, F. M. (1976). Adapting media attributes for effective learning. *Educational Technology, 16*, 7–13.

Fernald, G. M. (1943). *Remedial techniques in basic school subjects*. New York: McGraw-Hill.

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

Flick, L. B. (1993). The meanings of hands-on science. *Journal of Science Teacher Education, 4*, 1–8.

Glasson, G. E. (1989). The effects of hands-on and teacher demonstration laboratory methods on science achievement in relation to reasoning ability and prior knowledge. *Journal of Research in Science Teaching, 26*, 121–131.

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

Hatzikraniotis, E., Lefkos, J., Bisdikian, G., Psillos, D., Refanidis, J., & Vlahavas, J. (2001). An open learning environment for thermal phenomena. In G. Chapman (Ed.), *Proceedings of the fifth international conference on computer based learning in science (section ad3* (pp. 1–11). Brno, Czech Republic: Pedagogical Faculty of University of Ostrava.

Heron, P. R. L., & Meltzer, D. E. (2005). The future of physics education research: intellectual challenges and practical concerns. *American Journal of Physics, 73*, 390–394.

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

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

Jones, G. M., Andre, T., Superfine, R., & Taylor, R. (2003). Learning at the nano-scale: the impact of students' use of remote microscopy on concepts of viruses, scale, and microscopy. *Journal of Research in Science Teaching, 40*, 303–322.

Jones, G. M., Minogue, J., Tretter, T. R., Negishi, A., & Taylor, R. (2006). Haptic augmentation of science instruction: does touch matter? *Science Education, 90*, 111–123.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design; some food for thought. *Instructional Science, 38*, 105–134.

van Joolingen, W., & Zacharia, Z. C. (2009). Developments in inquiry learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S.

Barnes (Eds.), *Technology-enhanced learning: A kaleidosope view* (pp. 21–37). Dordrecht, The Netherlands: Springer.

Kirschner, P., & Huisman, W. (1998). Dry laboratories in science education; computer-based practical work. *International Journal of Science Education, 20*, 665–682.

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

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

Klatzky, R. L., Lederman, S. J., & Reed, C. (1987). There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General, 116*, 356–369.

Loomis, J. M., & Lederman, S. J. (1986). Tactual perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Cognitive processes and performance. Handbook of perception and human performances, Vol. 2* (pp. 1–42). New York: Wiley.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 147–158). Cambridge, UK: Cambridge University Press.

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

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

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

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction, 19*, 171–184.

van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*, 199–212.

Millar, S. (1999). Memory in touch. *Psicothema, 11*, 747–767.

Minogue, J., & Jones, M. G. (2009). Measuring the impact of haptic feedback using the SOLO taxonomy. *International Journal of Science Education, 31*, 1359–1378.

National Research Council. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.

National Science Teachers Association. (1999). Position statement on the use of computers in science education. Retrieved November 12, 2009, from the World Wide Web. http://www.nsta.org/positionstatement&psid=4.

Olympiou, G. (2006). Investigating the effects of physical and virtual manipulative experimentation on students' conceptual understanding in heat and temperature. Unpublished senior seminar thesis, Nicosia, Cyprus: University of Cyprus.

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

Reiner, M. (1999). Conceptual construction of fields through tactile interface. *Interactive Learning Environments, 7*, 31–55.

Resnick, M. (1998). Technologies for lifelong kindergarten. *Educational Technology Research and Development, 46*, 43–55.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*, 481–494.

Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: the development of the WHAT-IF test. *Studies in Educational Evaluation, 22*, 341–362.

Swaak, J., van Joolingen, W. R., & de Jong, T. (1998). Supporting simulation-based learning; the effects of model progression and assignments on definitional and intuitive knowledge. *Learning and Instruction, 8*, 235–252.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295–312.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

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

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

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

Zacharia, Z. C. (2003). Beliefs, attitudes, and intentions of science teachers regarding the educational use of computer simulations and inquiry-based experiments in physics. *Journal of Research in Science Teaching, 40*, 792–823.

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

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

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

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

Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching, 45*, 1021–1035.