International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Middle school students' learning of mechanics concepts through engagement in different sequences of physical and virtual experiments

Sarah Sullivan, Dana Gnesdilow, Sadhana Puntambekar & Jee-Seon Kim

To cite this article: Sarah Sullivan, Dana Gnesdilow, Sadhana Puntambekar & Jee-Seon Kim (2017): Middle school students' learning of mechanics concepts through engagement in different sequences of physical and virtual experiments, International Journal of Science Education, DOI: 10.1080/09500693.2017.1341668

To link to this article: http://dx.doi.org/10.1080/09500693.2017.1341668

Published online: 30 Jun 2017.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2017
https://doi.org/10.1080/09500693.2017.1341668

Routledge
Taylor & Francis Group

Check for updates

Middle school students' learning of mechanics concepts through engagement in different sequences of physical and virtual experiments

Sarah Sullivan[a], Dana Gnesdilow[b], Sadhana Puntambekar[b] and Jee-Seon Kim[b]

[a]School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; [b]Department of Educational Psychology, University of Wisconsin, Madison, WI, USA

ABSTRACT

Physical and virtual experimentation are thought to have different affordances for supporting students' learning. Research investigating the use of physical and virtual experiments to support students' learning has identified a variety of, sometimes conflicting, outcomes. Unanswered questions remain about how physical and virtual experiments may impact students' learning and for which contexts and content areas they may be most effective. Using a quasi-experimental design, we examined eighth grade students' ($N = 100$) learning of physics concepts related to pulleys depending on the sequence of physical and virtual labs they engaged in. Five classes of students were assigned to either the: *physical first* condition (PF) ($n = 55$), where students performed a physical pulley experiment and then performed the same experiment virtually, or *virtual first* condition (VF) ($n = 45$), with the opposite sequence. Repeated measures ANOVA's were conducted to examine how physical and virtual labs impacted students' learning of specific physics concepts. While we did not find clear-cut support that one sequence was better, we did find evidence that participating in virtual experiments may be more beneficial for learning certain physics concepts, such as work and mechanical advantage. Our findings support the idea that if time or physical materials are limited, using virtual experiments may help students understand work and mechanical advantage.

ARTICLE HISTORY
Received 1 May 2016
Accepted 8 June 2017

KEYWORDS
Physics; simulations; laboratory science; middle school

With advancing and increasingly affordable computing technologies, schools are no longer limited to physical materials like Bunsen burners and Erlenmeyer flasks to conduct experiments to teach science. Doing science can include dynamic animations and interactive simulations, enabling students to conduct both physical and virtual experiments. Now that use of animations and simulations in schools has become more commonplace, there is has been increased research interest in the differential benefits, and limitations, of physical and virtual experimentation (de Jong, Linn, & Zacharia, 2013; McKinney, 1997; Olympiou & Zacharia, 2012), with each mode of experimentation having their own benefits and drawbacks. For example, physical experiments offer the

CONTACT Sarah Sullivan sullivans@surgery.wisc.edu School of Medicine and Public Health, University of Wisconsin, 600 Highland Ave., H4/785B CSC, Madison WI 53792, USA

opportunity to investigate scientific concepts in real-world conditions but potentially miss important patterns in data due to measurement error. Alternatively, virtual experiments can facilitate an understanding of conceptual relationships under ideal conditions, but may create confusion about how these relationships express themselves in the real world. Therefore, the question of interest is not the simplistic one of which is best. Rather, it is for whom, under what conditions, and for what type of content virtual experimentation (or physical experimentation) is best suited – i.e. is students' learning from physical or virtual labs influenced by factors such as conceptual difficulty, the differences in the abstract/concrete dimension of the science concepts to be learned, or by the particular features that a physical or virtual experiment may offer?

Additionally, there are numerous plausible ways to sequence physical and virtual experiments in the science classroom. For example, students learning about pulley systems could set up and work with real pulleys to *feel* how the force needed to lift the load changes depending on the type of pulley system used and the amount of mechanical advantage (MA) it can provide. They could then go on to a virtual experimentation platform to replicate the physical experiments and investigate how measurement error might have influenced the results of the identical experiments in the real world to better understand the patterns in their data. Alternately, students might carry out idealised experiments on an inclined plane with no friction in the virtual world, then set up the exact same experiment in the real world where friction is omnipresent, and compare the results with a teacher's help to better understand the concept of friction. Is either of these sequences better than others in helping students to understand the underlying science at a deep level? One of the grand challenges in science education is to determine the optimal balance of physical and virtual investigations in science courses (Science, 2013). As such, an important question for educational research to address is: for the learning of what concepts and processes are physical experimentation, virtual experimentation, and different ways combining the two best suited?

The research reported in this paper is an attempt to answer some of the questions presented above. Specifically, this paper reports a study in which middle schools students learned topics in physical science by participating in physical and virtual pulley labs, as well as different sequences of the two types of experiments. Our aim was to explore if students' engagement in physical or virtual labs and different sequences of these labs supported students' learning of physics concepts related to simple machines, such as work and MA. For example, could it be the case that students benefit more from engaging in a physical experiment first to develop some perceptual grounding of concepts? Or, could it be the case that doing a virtual experiment first with idealised data (with no measurement error and zero friction conditions) would help students to focus on more important relationships and patterns between the different variables tested? The study reported represents a step towards answering these questions.

This paper is organised as follows. We will first discuss research on the affordances of physical and virtual labs as well as how students have learned from different sequences of these two types of labs. Then we will discuss the goals of this study and our research questions before moving into the details of the research and our results and conclusions.

Learning from physical and virtual experimentation

Both physical and virtual experiments are widely used in science classrooms to enable students to engage in science inquiry processes – such as designing experiments, collecting and analysing data, and using evidence to justify claims – all of which are emphasised in the *Next Generation Science Standards* (NGSS Lead States, 2013). While physical and virtual materials offer different affordances for learning, both environments can introduce students to the important conceptual and procedural knowledge of science and frame students' activities around important concepts in the domain (Hofstein & Lunetta, 2004). Both modes of exploration provide perceptual grounding for concepts that might otherwise be too abstract to be easily understood (Winn et al., 2006) and provide exposure to scientific experimentation and its corresponding skills (Hofstein & Lunetta, 2004).

Using physical experiments to teach science

Physical investigations that enable students to manipulate physical objects have been central to science learning (National Research Council, 2006) and are integral to many science curricula (e.g Krajcik, Blumenfeld, Marx, & Soloway, 1994). Only physical laboratory environments offer experiences to students that involve the manipulation of the actual items of a lab experiment, helping them to develop perceptual-motor skills (Olympiou & Zacharia, 2012). Physical labs enable students to experience science phenomena, collect data, and gain experience using measurement instruments. Activities that involve the building and testing of a physical model provide a natural environment for applying scientific principles and concepts, and provide ongoing feedback for students as they confront their understanding of scientific principles by trying to put them into practice. Model construction and testing entails the integration of several skills (Lehrer & Romberg, 1996), including analysis, synthesis, evaluation, and revision, which are often fragmented across the curriculum. Conducting physical experiments naturally includes measurement errors, which are inherently part of doing science and may be considered an affordance of conducting physical experiments (Olympiou & Zacharia, 2014), while computer simulations are often designed to avoid measurement error. Deep understanding of science concepts includes the knowledge of the types of measurement errors that exist in the domain and the ability to appropriately deal with them (Toth, Klahr, & Chen, 2000). Some researchers assert that for students to learn about the real-world phenomena themselves and to understand what science is really about requires engaging them in hands-on, 'live' experiments with real materials (Bell, 2004; Hofstein & Lunetta, 2004).

Further, the kinesthetic experiences associated with setting up experimental materials, 'running' an experiment, and taking measurements may have hitherto unexplored learning benefits. There is mounting evidence that our sensorimotor system affects cognition and learning in complex ways (Barsalou, 2009; Wilson, 2002). For example, according to theories of embodied cognition, our conceptual understanding is grounded in our experience in the world through metaphorical mappings (Lakoff & Johnson, 1980), and our conceptual processing is influenced by our actions, our use of physical manipulative, our bodily states, and the involvement of our sensorimotor brain regions (e.g. Glenberg, Brown, & Levin, 2007; Glenberg, Havas, Becker, & Rinck, 2005; Glenberg & Kaschak, 2002; Kontra, Lyons, Fischer, & Beilock, 2015).

However, Pouw, van Gog, and Paas (2014) reviewed several studies and discussed how physical experiences are not all created equal; specifically, 'perceptual and interactive richness in and of itself is not something that promotes learning, but is contextually dependent on the learning content being constituted in multimodal information' (p. 65). In other words, simply providing students with embodied experiences with physical materials in the real world may not be enough to ensure conceptual development, and there may be a multitude of factors that influence students' learning with physical materials.

Using virtual experiments to teach science

With the increasing use of computers in classrooms, dynamic visual representations in the form of virtual labs are being used in science learning and have several affordances. Addressing the limitations of physical laboratory experiments, Sadler, Whitney, Shore, and Deutsch (1999) argue that computer simulations can 'focus attention on formal variables, parameters, and frames of reference'. Simulations provide alternate representations that may not be possible with real materials (Zacharia & Anderson, 2003), such as trying out experiments in ideal situations (e.g. frictionless environments and elimination of measurement error). Simulations also allow the presentation of dynamically changing graphs and tables and enable students to test several 'what-if' scenarios. Furthermore, virtual experiments eliminate the risk of an experiment resulting in inconclusive or erroneous results (as discussed in Tatli & Ayas, 2013), as can be the case with physical experiments. Setting up simulations is generally less time consuming than setting up hands-on investigations, thereby allowing students more time to try out several conditions in their investigations and replicate their findings by running the experiments multiple times. Simulations can also combine multiple representations – verbal, numerical, pictorial, and graphical, facilitating a deeper understanding of underlying phenomena (Ainsworth, 1999, 2006). In addition, simulations allow students to perceive variables and conceptual relationships that are not directly observable in the physical environment (Bell, 2004; Snir, Smith, & Grosslight, 1995) as well as display visualisations of processes, which provide students with important feedback for learning (Zacharia & de Jong, 2014).

Simulations rely on models of real-world scenarios and, by necessity and definition, models strip away contextual elements to focus on specific phenomena. The models underlying simulations may be crafted to accommodate learners' prior knowledge and conceptions, support learning trajectories, and build bridges between contextualised examples and formal abstractions (White, 1993). However, when students interact with simulations they are often interacting with simplified models, not with the more complex, contextualised, and messy real world the models represent (Bell, 2004; Hofstein & Lunetta, 2004). Thus, while simulations can help control parameters, simulated results may not adequately represent the real world.

Virtual versus physical experiments for learning

There are obvious advantages to carrying out experimentation-driven inquiry learning in virtual laboratories as opposed to physical ones. Virtual experiments allow unique testing conditions, such as zero friction or gravity that are all but impossible to create in a real school laboratory (Zacharia & Anderson, 2003). Simulations can also make entities, variables, and relationships visible that cannot be seen in a real lab, such as force vectors or virus particles (Bell, 2004; Snir et al., 1995; Winn et al., 2006). Real experimentation can be messy, expensive, and is not always the best use of students' time. Students may spill chemicals in a chemistry lab or stumble in setting up and threading complex, multi-pulley systems in a physics lab. Further, they may spend more time on procedural issues rather than productive learning conversations as they run into process-related problems when setting up complex physical experiments (Zacharia & de Jong, 2014). Such issues eat into the already limited time available for lab work and may hinder students' learning. Then there is the issue of measurement errors. Instrument readings may be inaccurate due to poor calibration, student errors, or physical limitations.

When directly comparing physical and virtual experimentation, several studies have found no significant and consistent differences between learning from simulations and physical laboratories (Ma & Nickerson, 2006; Triona & Klahr, 2003; Zacharia & Constantinou, 2008; Zacharia & Olympiou, 2011). However, there have also been instances where the use of virtual laboratories has better supported students' learning than physical laboratories (Finkelstein et al., 2005; Zacharia, 2007; Zacharia, Olympiou, & Papaevripidou, 2008). In one study, Pyatt and Sims (2012) found similar results as those mentioned above: one investigation showed no significant difference in students' performance through participating in a physical or virtual experiment and a second investigation showed that students' performance using a virtual experiment was significantly better. However, these researchers explained the difference in the conceptual understanding between the two groups in the second investigation might have been due to differences in the quality of data that the students were able to gather, since the data gathered in the virtual condition provided a more accurate picture of the underlying science. Alternatively, a qualitative study found that physical experiments were more beneficial and less time consuming than using computer simulations for pre-service teachers' development of ideas and explanations about the way things work, or theories-in-action, in physics (Marshall & Young, 2006).

Combining virtual and physical experiments

Though early research on physical and virtual science investigations has treated physical experimentation and computer simulations as competing methods in science classrooms (Jaakkola & Nurmi, 2008), it is, of course, not always a choice of one or the other. It is often feasible to combine and sequence the two modes of experimentation in a multitude of ways. Recent research is beginning to support the idea that virtual and physical experimentation each have unique affordances for learning (Olympiou & Zacharia, 2012, 2014; Zacharia & de Jong, 2014) and therefore need to be combined for optimal learning. Physical labs may be needed to restore the contextual elements missing in virtual experiments to enable students to situate their experience and learning, and thus facilitate transfer of what they have learned in the simulation to the real world. Hence, learners may ground the simulation in the physical world, both perceptually and experientially. Over the past few decades, several research studies have attempted to investigate the value of using physical laboratory environments and virtual laboratory environments within science classrooms to support student learning (Olympiou & Zacharia, 2012) and several researchers have advocated for conscientiously combining hands-on and virtual experiments to help facilitate students' conceptual understanding and their knowledge of how science is done (Chinn & Malhotra, 2002; Snir et al., 1995).

When considering combinations of physical and virtual experiments in comparison to each by itself, several studies have found that combined physical/virtual experiments can be more beneficial for learning than either form alone (Jaakkola & Nurmi, 2008; Olympiou & Zacharia, 2014; Zacharia, 2007; Zacharia et al., 2008), though that is not always the case (Zacharia & Olympiou, 2011). In a study where Zacharia and de Jong (2014) examined five different virtual/physical sequences with high school students during an electrical circuits unit, they found that not all physical and virtual experimental sequences were equally effective in helping students to develop conceptual understanding. In particular, they found that students who performed a virtual experiment at a key point in the curriculum, where they were able to get feedback about electrical circuits through the simulation, did better than students who performed physical experiments for that same part of the experiment, regardless of whether they did physical or virtual experiments before or during the remainder of the curriculum. They stated that it 'is not a matter of which mode of experimentation that should proceed the other … , but rather a matter of when a mode of experimentation, along with its affordances, really contributes to learning' (p. 147).

Within the specific context of learning with pulleys, a study looking at undergraduates' learning from physical and virtual experiments found that students' performance on a post-test was related to both the concept being learned and whether the post-test was immediate or delayed. This study found that both physical and virtual experiments facilitated undergraduates' learning about the concepts of force and MA equally. Nevertheless, virtual experiments were found to significantly better support students' learning about work and energy (Chini, Madsen, Gire, Rebello & Puntambekar, 2012). However, these results have not been consistent across studies. An experiment examining sequencing of virtual and physical pulley experiments with sixth grade students found that students learned more if they conducted a virtual experiment after a physical experiment, but that this finding may have been related to particular concepts that the experiments were designed to highlight. For example, students seemed to learn about the concepts of MA and potential energy equally well from both physical and virtual experiments. However, students appeared to learn more effectively about the concept of work if they conducted a physical experiment first and, alternatively, learned more about force from participating in a virtual experiment first (Gire et al., 2010). Given the paucity of research and conflicting results on the learning of physics concepts related to this commonly taught simple machine, further research in this area needs to be undertaken to better understand the benefits and affordances of both types of experimentation, how they might be effectively combined in particular contexts or content areas, and who might benefit most from them.

Goals of the present study

As mentioned earlier, the present study was conducted to explore the intricacies that may be involved in middle school students' learning of science concepts through participating in different sequences of virtual and physical experiments. While there are many studies examining how undergraduate students use physical or virtual experiments, fewer studies have looked at middle school students' learning from these different types of experiments.

We were interested in understanding if physical or virtual labs and different sequences of the two might differentially affect middle students' learning of particular concepts. For example, based on an embodied view of learning, one might predict that experimenting with pulleys in the real world would help students to develop a better conception of applied force, since they could actually *feel* the difference in the amount of force needed in moving the load. On the other hand, we might expect students to learn the concept of work better using a simulation, where the data collection is less messy, enabling students to better see patterns and relationships in the data they collect. This may be particularly true given that pulleys are simple machines that are often challenging for younger students to set up on their own, making it challenging for them to test multiple set-ups in a single class period. To this end, we were interested in exploring the following questions:

(1) Do physical labs help students to learn the physics of simple machines concepts better than virtual labs (or vice versa), as evidenced by performance on conceptual assessments?
(2) Do different sequences of the two types of labs differentially affect middle school students' learning of physics concepts within the topic of pulleys?

Methods

Context of the research

One hundred 8th grade students and one teacher from a mid-western, public, middle school in the USA participated in this research during the 2010–2011 school year. The school was located on the urban fringe of a mid-sized city. In 2010, the student population consisted of predominately (94.3%) white, middle class students, with only 5.5% being categorised as economically disadvantaged. A total of five 8th grade classes were involved in this study. While each class was randomly assigned to one of two conditions described below, this study is considered a quasi-experimental design, since students within each classroom could not be randomly assigned to treatment conditions.

The collaborating teacher who implemented the CoMPASS curriculum (described below) and participated in the research taught five sections of eighth grade science. Prior to this study, she had five years of experience teaching science (using CoMPASS during two of these years) and a bachelor's degree in elementary education with certification to teach science in grades 1–9.

The CoMPASS Work and Energy unit used during this study, an inquiry and design-based curriculum, takes students about eight weeks to complete. At the beginning of the unit, students were presented with a design challenge to create a compound machine that would help one of the teachers in the school, who had a wrist injury, to lift heavy items to help her accomplish daily tasks without the help of others. Their goal was to decrease the amount of force she would need to exert to lift heavy things. The Work and Energy unit consisted of six mini challenges; one challenge for each of six simple machines so students could participate in cycles of inquiry and learn about science concepts over time.

Students worked in the same collaborative groups of approximately four students throughout the entire unit to ask questions, conduct research and experiments, analyse findings, and identify and communicate science relationships relevant to solving their challenge. They also used the CoMPASS e-textbook to learn about physics concepts to prepare them for their design challenges. All of the activities were designed to help students learn physical science concepts and relationships such as applied force, work, MA, and potential and kinetic energy. Students were provided with both physical materials for all of the challenges and virtual simulations for the inclined plane and pulley aspects of the unit to conduct experiments. During the physical pulley experiment, students set up fixed, movable, and double and triple compound pulley systems, lifting the same load with each configuration.

Because pulleys are difficult for students to set up and take accurate measurements, the pulley simulation allowed the students to explore a similar variety of pulley configurations with a load of the same mass as in the physical experiment. The simulation environment that students used to conduct their virtual experiments was called the Virtual Physics System (ViPS) (Myneni, Narayanan, Rebello, Rouinfar, & Pumtambekar, 2013) (see Figure 1 below). ViPS is an intelligent tutoring system that provides students with hints for solving problems and virtually designing pulley experiments. The simulation afforded the ability for students to change several design parameters, such as type of pulley system, length of rope, size of pulleys, and friction, and quickly and accurately receive the output measurements for their dependent variables, such as applied force, work, MA, and potential energy. For both the physical and virtual pulley experiments, students recorded

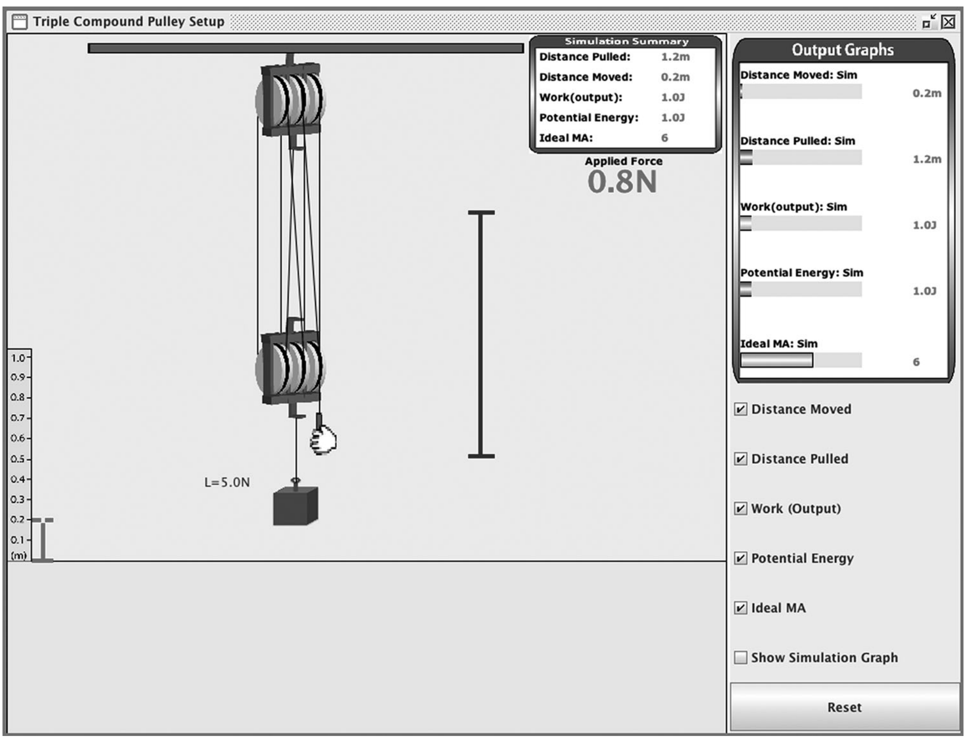


Figure 1. Screen shot of the ViPS pulley simulation environment.

data in data charts in their student notebooks, in order to allow them to make connections among patterns of data in the variables involved in the experiments.

Design and procedure

This study occurred as a part of students' regular science classes during the pulley unit, the final part of the Work and Energy CoMPASS unit. Students spent approximately 16 forty-five minute class periods, or 12 hours, working on the pulley unit. Each class was assigned to one of two conditions. The two conditions, (a) *physical first* (PF) and (b) *virtual first* (VF), were examined to explore whether the sequence of physical and virtual experiments had an impact on students' learning outcomes.

In both conditions, students were asked to solve the pulley challenge: to determine how to increase the amount of MA, reduce the amount of applied force, and reduce the work and energy required to lift a large and heavy statue of their school mascot onto a pedestal using pulleys. In the physical condition, students experimented with a scaled down model of this scenario using pulleys, string, and a water bottle as the load representing the mascot statue. In the PF condition, students took a pulley content knowledge and relationships pretest, conducted physical system height and energy pulley experiment to explore physics concepts, and then took the pulley content knowledge and relationships mid-test. Subsequently, they completed the same pulley system and height and energy experiment using a virtual simulation and then took the pulley content knowledge and relationships post-test. This order was reversed in the VF condition. In both cases, students took the same pre, mid, and post-tests. Students in both conditions also engaged in generating questions and finding information from the e-textbook before doing experiments. All five 8th grade classes spent the same amount of time participating in the experiments and unit regardless of the treatment condition they were in. Three classes participated in the PF condition and two classes participated in the VF condition. Although classes were randomly assigned, we checked with the teacher to makes sure the relative abilities of the

Table 1. Sequence of the study by PF and VF conditions.

	Pretest	Mid-test		Post-test
Condition 1 PF $N = 55$	• Introduction to pulley challenge • Anticipation guide • Question generation and sharing • CoMPASS exploration	Physical system and physical height and energy pulley experiments	Second CoMPASS exploration	Virtual system and virtual height and energy pulley experiments
Condition 2 VF $N = 45$	• Introduction to pulley challenge • Anticipation guide and sharing • Question generation • CoMPASS exploration	Virtual system and virtual height and energy pulley experiments	Second CoMPASS exploration	Physical system and physical height and energy pulley experiments

PF, physical first condition; VF, virtual first condition.

classes were distributed evenly so as not to advantage one treatment condition over the other. There were five school days between the pre and the mid-tests and four school days between the mid and post-tests. Table 1 shows the number of students in each condition, the sequence of activities, and the points at which the tests were taken for both conditions.

Measures

We constructed a test consisting of 16 multiple-choice and 13 open-ended questions (29 total questions). The maximum possible score on the test was 44 points (16 points for the multiple-choice and 28 points for the open-ended questions). Several experts in the fields of education, physics, and test construction were involved in the design and revision of all of the pulley test questions to ensure content validity. Questions were designed to assess (i) students' knowledge of physics concepts related to pulleys, including *applied force, MA, friction, work* and *potential energy* and (ii) relationships between these concepts under different conditions, such as the height an object is lifted and potential energy it has at the top. The same test was used as the pre, mid, and post-test during the experiment. Cronbach's alpha for the pre, mid, and post-tests was .64, .78, and .77, respectively. See the Appendix for the test questions, with the concept targeted in each question highlighted.

The multiple-choice questions presented four possible choices and students were instructed to choose only one letter to indicate the answer for each question. Students' responses to the open-ended questions were scored using rubrics that ranged from either 0 to 2 or 0 to 3 points, depending on the complexity of the question. For some questions, we identified only two levels of responses. For other questions that were more complex, students' responses could be categorised into three levels. Students were given points based on whether their responses were incorrect, partial with a lay explanation, or complete with a scientific explanation of relationships. For the 2-point rubric, incorrect responses received a 0, a response that mentioned a single concept correctly with partial explanation or which used lay terminology to refer to concepts received a 1, and a response that described the relationships and causal mechanisms among concepts in scientific terms received a 2. For the 3-point rubric, incorrect responses received a 0, a response that mentioned a single concept correctly with partial explanation or which used lay terminology to refer to concepts received a 1, a response that described the relationships and causal mechanisms among concepts in scientific terms received a 2, and a response that went beyond describing the focal relationships and causal mechanisms to incorporate other physics concepts and their impact on the concepts of focus received a 3. In order to establish inter-rater reliability, two raters coded 10% of the open-ended responses on the pulley test. After using this set of questions to discuss any disagreements, the raters coded a second 10% of the responses and achieved an agreement of 90%. Disagreements were resolved through discussion. The raters then independently coded the remaining responses.

Analyses

To examine if participation in a physical or virtual experiment or a particular treatment sequence, PF or VF, would influence students' learning of physics content related to pulley systems, we broke the pulley test questions up into four subsets: (1) applied force, (2) distance, (3) work, and (4) MA. The content question subsets were created by two independent raters assigning each question in the pulley test to one of the main four content subcategories. A 97% inter-rater agreement was achieved after one round of coding of all questions. The categorisation of the one question where there was a discrepancy between raters was resolved by discussion. The number of questions and total possible score per category are as follows: 8 applied force questions with a total of 11 possible points, 5 distance related questions with a total of 7 possible points, 8 questions about work with a total of 12 possible points, and 8 questions about MA with a total of 14 possible points.

Results

Effect of treatment on students' performance on content subsets of questions

We conducted a 2 (treatment: PF, VF) × 3 (test-time: pre, mid, post subset scores) mixed ANOVA to compare the effect of treatment on students' learning of specific subsets of questions on the pulley test. The dependent measures were students' scores on four subsets of questions: (1) applied force, (2) distance, (3) work, and (4) MA, with the same test questions being used to assess learning at the pre, mid, and post-test times. Test-time was the within subjects factor and treatment (PF or VF) was the between subjects factor. When using $\alpha = .05$, we did not find differences between PF and VF students' learning in the applied force and distance subsets, but did find significant differences in their learning of the concepts of work and MA. See Table 2 for subset descriptive statistics and Figure 2 for plots of students' performance over time by treatment on each of the subsets. To examine the order effects as well as initial and additional learning from the two experimentations, we conducted a series of simple effect contrasts for testing the mean differences between the mid-test scores while controlling for the pretest scores, the post-test scores while controlling for the mid-test scores, and the post-test scores while controlling for the pretest scores. When we found significant differences in concept learning, we conducted follow-up pairwise comparisons to better understand these differences. In order to control for type I error, we used Holm's sequential Bonferroni approach (Holm, 1979; Hommel, 1988). In this approach, one divides the established alpha, in our case .05, by the number of follow-up pairwise comparisons performed in a sequential manner. To do this, we obtained the p-values of all of the pairwise comparisons within each family of tests and ordered them from smallest to largest. The pairwise comparison with the smallest p-value is tested against the established alpha divide by the total number of comparisons in the family. If this comparison is found to be significant, the pairwise comparison with the second smallest p-value is tested against the established alpha divided by one less comparison. This continues in the same manner until a non-significant result is found.

Applied force subset

In running the analyses for the applied force subset, there was no significant interaction between test-time and treatment, Wilks' $\Lambda = .943$, $F(2,97) = 2.91$, $p = .059$, partial $\eta^2 =$

Table 2. Means and standard deviation for PF and VF students' scores on pulley test subsets.

			95% Confidence interval			95% Confidence interval			95% Confidence interval	
Test subset	Treatment	Pre	Lower bound	Upper bound	Mid	Lower bound	Upper bound	Post	Lower bound	Upper bound
Applied force (11 possible pts.)	PF	4.38(1.33)	4.02	4.74	5.62(1.43)	5.19	6.04	6.07(2.27)	5.48	6.67
	VF	4.82(1.42)	4.42	5.23	5.49(1.77)	5.02	5.96	6.51(2.18)	5.85	7.17
Distance (7 possible pts.)	PF	2.67(1.04)	2.42	2.93	3.75(1.60)	3.30	4.19	3.96(1.59)	3.53	4.40
	VF	2.44(0.87)	2.16	2.73	3.42(1.71)	2.93	3.91	3.84(1.67)	3.37	4.32
Work (12 possible pts.)	PF	4.20(2.22)	3.59	4.81	3.93(2.36)	3.24	4.61	5.15(2.98)	4.33	5.96
	VF	4.04(2.33)	3.37	4.72	5.29(2.80)	4.53	6.05	5.36(3.14)	4.45	6.26
MA (14 possible pts.)	PF	3.11(1.75)	2.66	3.56	5.51(2.67)	4.72	6.30	6.58(3.25)	5.74	7.42
	VF	2.60(1.62)	2.10	3.10	6.82(3.25)	5.95	7.69	7.29(3.01)	6.36	8.22

PF, physical first condition, VF, virtual first condition, MA, mechanical advantage.

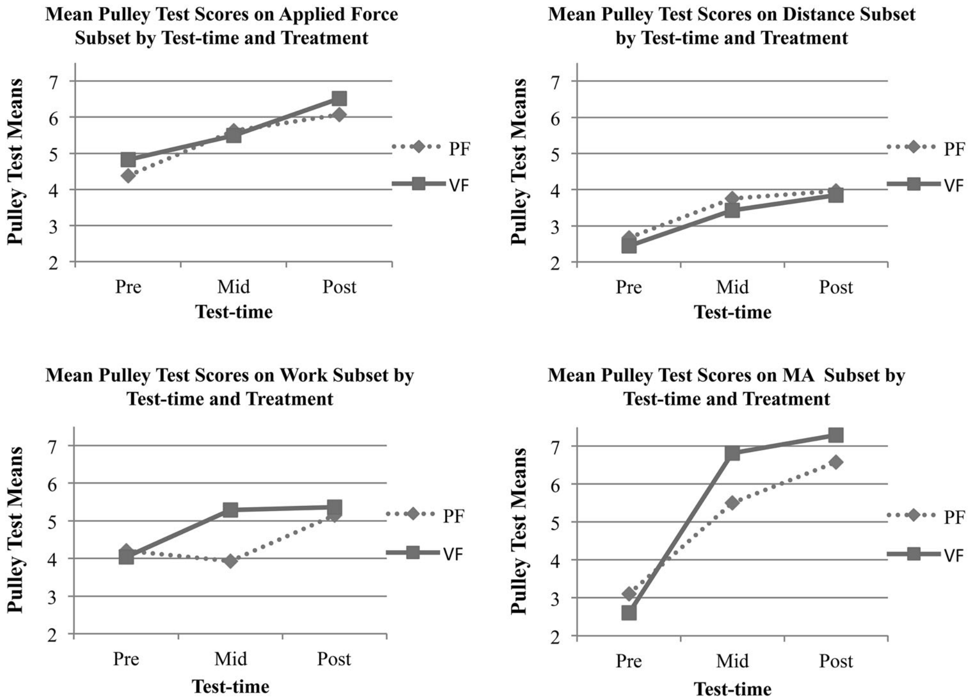


Figure 2. Pre, mid, and post-test scores by condition on the applied force, distance, work, and MA subsets of the pulley test.

0.057, indicating that students' learning of applied force concepts was similar among the treatment two groups over time.

However, we found a significant main effect for test-time, Wilks' $\Lambda = .570$, $F(2,97) = 36.61$, $p < .001$, partial $\eta^2 = 1.00$, on the applied force subset. To further understand this difference, we conducted two follow-up pairwise comparisons. We found that students made significant gains from pre to mid ($t_{98} = 6.39$, $p < .001$, $d = 0.64$) and mid to post-test ($t_{98} = 4.64$, $p < .001$, $d = 0.85$) for the applied force questions. This indicates that students made significant gains in learning applied force concepts over the entire sequence of pulley activities.

Distance subset

We found a similar trend for the analyses run on the distance subset. Again, there was no significant interaction between test-time and treatment, Wilks' $\Lambda = .997$, $F(2,97) = .150$, $p = .861$, partial $\eta^2 = 0.072$, indicating that students' learning of applied force concepts was similar among the two treatment groups over time.

We did, however, find a significant main effect for test-time, Wilks' $\Lambda = .540$, $F(2,97) = 41.32$, $p < .001$, partial $\eta^2 = 1.00$, on the distance subset. Two follow-up pairwise comparisons showed that students only made significant gains from pre to mid-test ($t_{\mathrm{df}} = 6.37$, $p < .001$, $d = 0.64$) and did not make significant gains from mid to post-test ($t_{98} = 1.72$, $p = .088$) for the distance subset questions. Thus, no matter what treatment group students were in, they did not make significant gains from participating in the second experiment.

Work subset

Unlike with the applied force and distance subsets, we did find a significant interaction between test-time and treatment for the work subset. The results indicated that students in the PF and VF groups performed significantly differently on the subset of work questions over time, Wilks' $\Lambda = .924$, $F(2,97) = 3.40$, $p = .021$, partial $\eta^2 = 0.076$).

To follow up on this finding, we ran four pairwise comparisons. We found that the PF students did not make significant gains from pre to mid-test ($t_{98} = -0.74$, $p = .460$). In fact, PF students' scores actually decreased slightly from pre to mid-test. But, they did make significant gains from mid to post-test on the work subset of the pulley test, after participating in the virtual experiment ($t_{98} = 2.63$, $p = .006$, $d = 0.38$). VF students showed an opposite pattern; they made significant gains from pre to mid-test ($t_{98} = 3.06$, $p = .003$, $d = 0.46$), but did not make significant gains from mid to post on the work subset of the pulley test ($t_{98} = 0.14$, $p = .890$). These results indicated that both groups of students only made significant learning gains after completing the virtual experiment, but did not make significant gains from participating in the physical labs.

Three follow-up pairwise comparisons, revealed that the VF group did significantly better than the PF group on the mid-test ($t_{\mathrm{df}} = -2.64$, $p = .010$, $d = 0.26$), but did not do significantly better on the pre ($t_{98} = 0.342$, $p = .734$) or post-test ($t_{98} = 0.342$, $p = .733$). So, students in the VF condition scored significantly better than PF students on the mid-test, but the PF students 'caught up' to the VF students after participating in the simulation experiments by the end of the unit.

MA subset

As with the work subset, we did find a significant interaction between test-time and treatment for the MA subset, Wilks' $\Lambda = .926$, $F(2,97) = 3.86$, $p = .024$, partial $\eta^2 = 0.074$, indicating that students' learning of MA concepts significantly differed between the PF and VF groups over time. We ran four pairwise comparisons to explore how the two groups performed differently on the MA subset over time. We found that the PF students made significant gains from pre to mid-test ($t_{98} = 5.29$, $p < .001$, $d = 0.71$). However, PF students' did not make significant gains from mid to post-test on the MA subset of the pulley test ($t_{98} = 2.09$, $p = .039$). VF students showed a similar pattern as the PF students on the MA subset. They also made significant gains from pre to mid-test ($t_{98} = 8.41$, $p < .001$, $d = 1.25$), but not from mid to post-test ($t_{98} = 0.82$, $p = .413$). These results indicated that both groups of students did most of their learning by engaging in the first experiment, regardless of the treatment.

Follow-up pairwise comparisons showed that there were no significant differences between the PF and VF students' scores on the pretest ($t_{98} = 1.49$, $p = .137$), mid-test ($t_{98} = 2.22$, $p = .029$), and post-test ($t_{98} = -1.12$, $p = .265$). While the difference between scores on the mid-test was not significant between the two treatment groups when using our stringent Holm's sequential Bonferroni approach to control for type I error rejection region, it is important to note that the PF group's average MA subset score was lower on the post-test than the VF students' average score on the mid-test (see Figure 2).

Discussion

One of the grand challenges for science education is to discern the best practices for using physical and virtual investigations in the science classroom. Thus far, the findings in this area have not been clear-cut and there is a still a great deal of work to be done. Developing a better understanding of these issues is a high priority as both physical and virtual experiments are ubiquitous in science classrooms for allowing students to learn both conceptual and procedural knowledge. However, understanding when, how, and under which context each form or sequence of physical and virtual experiments should be used is less obvious.

To gain insights into the best ways to use physical and virtual labs, we examined the effect of the sequence of physical and virtual experiments and the effect of individual physical and virtual labs on learning specific content in physics. While we did not find clear-cut support that one sequence was better than another in helping students to learn physics content, we did find evidence that participating in virtual experiments may be more beneficial for learning certain physics concepts such as work and MA than participating in physical experiments.

Research suggests that the combination of the two forms of experimentation, and the unique affordances they offer, will help students to learn better than doing either a physical or virtual experiment alone (Jaakkola & Nurmi, 2008; Olympiou & Zacharia, 2014; Zacharia, 2007; Zacharia et al., 2008). In exploring how the sequence of physical and virtual experiments impacted students' learning of particular physics concepts, we did find some differences. Although we did not find any significant differences in performance on the pulley tests between the PF and VF groups for content questions related to applied force and distance, we found that engaging students in the VF condition was more beneficial in helping them to learn the concepts of work and MA.

Based on an embodied cognition perspective that physical interactions in the environment influence cognition and learning (Barsalou, 2009; Goldstone, Landy, & Son, 2008; Kontra et al., 2015), one might have predicted that students' kinesthetic experiences with physically feeling how heavy or light a load is and pulling the string for longer distances when using a pulley would have helped students in the PF condition to better understand the concepts of applied force and distance. However, we did not find any significant differences between the PF and VF groups in their learning of these concepts. Perhaps these findings can be explained by the fact that the load that was being lifted, a small plastic bottle filled with water, was not very heavy to lift unaided by a simple machine in the first place. Maybe, if the object being lifted was heavier, or impossible to lift with a fixed pulley, but possible with other pulley systems, we may have found greater learning benefits from participating in the physical experiment. It is conceivable that in the early stages of learning, our physical experiences need to be more differentiated than was the case in our physical experiments to foster students' learning of the concept of applied force. Doing both experiments in either order seemed to help students learn about the concept of applied force, because students' test scores improved after each experiment, no matter the order.

In terms of learning about the concept of distance, we found that no matter the treatment condition, students did the bulk of their learning about the concept during their first experiment, as students in both treatment groups did not do significantly better on the post-test than the mid-test on the distance subset. Perhaps it is the case that the concept of distance is simply easy to understand on a conceptual level based on a multitude of prior, real-life and mathematical experiences. This may be why we did not find differences in students' learning of the concept depending on whether they simply saw larger distances in the simulation output or had the experience of pulling a longer distance while doing the physical experiment. However, students' average score on the pulley mid and post-tests were well below the maximum score possible, including scores for the distance subset. This finding may indicate that how the concept of distance relates to the physics of simple machines, including relationships to work and applied force, was difficult for students to grasp from physical and virtual experiments or a combination of the two.

We found that VF students made significant gains from the pre to mid-test, but not from the mid to post-test and that the PF students had an opposite result on the work subset. This finding indicates that students only made gains after participating in the virtual experiment, and the physical experiment was not very effective in helping students to understand the concept of work. In fact, students in the PF group actually did worse on the mid-test than the pretest, though not significantly so. In addition, we found that students in the VF group scored significantly higher on the mid-test than the PF students, but the PF students 'caught up' after participating in the virtual experiment, since there were no significant differences between the groups' scores on the post-test. In looking at the data collected during the physical and virtual experiments, we found that students' data for work during the physical experiment were quite messy and did not show idealised patterns in the data for work. This was most likely due to measurement error in data collection. There is a certain amount of estimation that takes place in reading a spring scale and measuring the lengthy pieces of pulley string that is necessary for calculating work. In addition, sometimes students do not thread their pulleys properly, which causes the string to get stuck causing the applied force values to go up as they pull. These issues potentially affect students' work calculations and, hence, cause the patterns for work to be conceptually inaccurate representations of the trade-off between applied force and distance. On the other hand, the data for work collected in the virtual simulation showed very clear and scientifically accurate patterns.

Along these same lines, clear patterns in data charts in students' notebooks may have also influenced students' learning of MA concepts. However, our follow-up pairwise comparisons were not significant and indicated a similar pattern of performance on the pulley tests over time between the two groups. Overall, the students did most of their learning by engaging in the first experiment, regardless of the treatment. However, the students in the VF group had a higher average score at the mid-test than the PF students had on the post-test. This may lend support for the idea that engaging in the virtual experiment after engaging in the physical one was beneficial for the students in the PF treatment, whereas the VF group did not benefit as much from engaging in the physical experiment after participating in the virtual one. These somewhat tenuous findings could, again, be explained by the idea that students' calculations for MA in the physical experiment did not show an idealised pattern in the data so that the concept of MA could readily be seen in relationship to other variables and concepts in the experiment. Thus, the patterns in the data that the students could use to learn more about the concept of MA were idealised in the virtual experiment and may have supported a better conceptual grounding of the idea.

With all other things being equal, it seems as though students benefit from idealised data that clearly shows the patterns of the underlying science relationships – such as in the case for work and MA described above. So, our findings lend support for the idea that if there is limited time, having students engage in a virtual experiment, where data collection is less messy, might be more effective than a physical one for learning about physics within the topic of pulleys. However, if there is no access to a virtual experiment, our results indicate that a physical experiment may be just as beneficial for every concept except for work. Educators should be mindful of issues with learning about work in physical environments with pulleys due to measurement error and the inability to see idealised patterns in their data, since there is always some friction in the real world. Given that this simple machine is a common topic used to teach about physics concepts in middle school science curricula, the results of this study have practical applications for science teachers. Our findings support the idea that if you can only do one experiment, do a virtual one. Students appeared to benefit from idealised data for developing an understanding the concept of work, and based on our findings, seemed to benefit equally well from either modality for learning about other physics concepts within the topic of pulleys.

Limitations and future research

Our results suggest that virtual experiments might be more beneficial for supporting students' understanding of concepts such as work or MA, and perhaps other concepts where conceptual patterns are more readily seen when measurement error is not a factor. While this study adds to the research on the benefits of physical and virtual experiments to foster students' learning by identifying that different modes of experimentation may be more effective in helping students to learn particular science concepts and that engaging students in virtual experiments may be more beneficial than physical in some contexts, there are limitations to this study.

There were some methodological challenges in this study. First of all, the subsets of content questions were determined by raters based on the intent of the test questions. However, a factor analysis did not reveal each of the subsets to be completely independent factors. This may be because of the multiple relationships among concepts, and thus, it may be difficult to separate them into independent subsets testing different concepts. It may also be because although the four subsets of the test reflect different aspects of the pulley experimentation, such as applied force and distance, these subsets do not directly correspond to different constructs in physics. Rather, the test is designed to measure students' common knowledge and understanding of physics related to pulleys. Secondly, some of our test items have a two-tier structure, using both multiple-choice and open-ended questions to test students' understandings of concepts and their relationships. Even though these questions may be testing different levels of knowledge, doing separate analyses with multiple-choice and open-ended questions revealed no differences in results. Therefore, the analyses in this study were done using the overall combination of these items for the concept subsets.

One important thing to note is that while students in both the PF and VF treatment groups did make significant gains from participating in the experiments, they did not max out the possible score of the pulley test within each subset of interest. Perhaps this was due to the nature of the experiments or the difficulty of our test. Additionally, it may be the case that setting up pulleys is simply a difficult process that is distracting and frustrating to students. Thus, they are concentrating on the procedure of setting things up rather than thinking and talking about the underlying science, as was found by Zacharia and de Jong (2014). We plan on examining students' dialogue as they participated in the PF and VF sequences to better understand the kinds of discourse the students were involved in as they explored physics concepts related to pulleys. Perhaps students engaged in particular kinds of talk depending on if they were involved in a virtual or physical experiment or participating in a physical experiment after participating in a virtual one (or vice versa). The information gathered from looking at this process data will be a productive next step in understanding the results of this study and building upon the prior research in this area in order to provide additional information regarding the timing of physical and virtual experiments when teaching physics concepts to middle school students in the context of the pulley as a common simple machine.

Acknowledgements

We thank our collaborators, Hari Narayanan at Auburn University and Sanjay Rebello at Purdue University, for their input into the development of this study. We also thank the teachers and students who participated in and made this study possible.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is supported in part by U.S. Department of Education, Institute of Education Sciences, Grant No. R305A080507 and the U.S. National Science Foundation Education Core Research Program, Grant No. DRL-1431904.

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers and Education*, *33*(2–3), 131–152.

Ainsworth, S. (2006). Deft: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*(3), 183–198.

Barsalou, L. (2009). Situating concepts. In M. Aydede & P. Robbins (Eds.), *The Cambridge handbook of situated cognition* (pp. 236–263). Cambridge: Cambridge University Press.

Bell, P. (2004, July). *The school science laboratory: Considerations of learning, technology, and scientific practice*. Paper presented at the meeting of High School Science Laboratories: Role and Vision, National Academy of Sciences, Washington, DC.

Chini, J. J., Madsen, A., Gire, E, Rebello, N. S., & Puntambekar, S. (2012). Exploration of factors that affect the comparative effectiveness of physical and virtual manipulatives in an undergraduate laboratory. *Physical Review Special Topics-Physics Education Research*, *8*(1), 010–113.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218.

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, *340*(6130), 305–308. doi:10.1126/science.1230579

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., … LeMaster, R. (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), 010103, 1–8.

Gire, E., Carmichael, A., Chini, J. J,, Rouinfar, A., Rebello, S., Smith, G., & Puntambekar, S. (2010). The effects of physical and virtual manipulatives on students' conceptual learning about pulleys.

In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1* (pp. 937–943). Chicago, Illinois: International Society of the Learning Sciences.

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, *32*(3), 389–399.

Glenberg, A., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states: The case for emotion. In D. Pecher & R. A. Zwaan (Eds.), *The grounding of cognition: The role of perception and action in memory, language, and thinking* (pp. 115–128). Cambridge: Cambridge University Press.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Goldstone, R., Landy, D., & Son, J. Y. (2008). A well-grounded education: The role of perception in science and mathematics. In M. DeVega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment and meaning* (pp. 327–355). Oxford: Oxford University Press.

Hofstein, A., & Lunetta, V. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28–54.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, *75*(2), 383–386.

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

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science*, *26*(6), 737–749.

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal*, *94*(5), 483–497.

Lakoff, G. J., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, *14*(1), 69–108.

Ma, J., & Nickerson, J. V. (2006). Hands-on, simulated, and remote laboratories: A comparative literature review. *ACM Computing Surveys (CSUR)*, *38*(3), 7–es.

Marshall, J. A., & Young, E. S. (2006). Preservice teachers' theory development in physical and simulated environments. *Journal of Research in Science Teaching*, *43*(9), 907–937.

McKinney, W. J. (1997). The educational use of computer based science simulations: Some lessons from the philosophy of science. *Science & Education*, *6*(6), 591–603.

Myneni, L. S., Narayanan, N. H., Rebello, S., Rouinfar, A., & Pumtambekar, S. (2013). An interactive and intelligent learning system for physics education. *IEEE Transactions on Learning Technologies*, *6*(3), 228–239.

National Research Council. (2006). *America's lab report: Investigations in high school science*. Committee on High School Science Laboratories: Role and Vision (S. R. Singer, M. L. Hilton, & H. A. Schweingruber, eds.). Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education (pp. 75–115). Washington, DC: The National Academy Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

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

Olympiou, G., & Zacharia, Z. C. (2014). Blending physical and virtual manipulatives in physics laboratory experimentation. In C. Bruguière, A. Tiberghien, & P. Clément (Eds.), *Topics and trends in current science education* (pp. 419–433). Dordrecht: Springer.

Pouw, W. T., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51–72.

Pyatt, K., & Sims, R. (2012). Virtual and physical experimentation in inquiry-based science labs: Attitudes, performance and access. *Journal of Science Education and Technology*, *21*(1), 133–147.
Sadler, P. M., Whitney, C. A., Shore, L., & Deutsch, F. (1999). Visualization and representation of physical systems: Wavemaker as an aid to conceptualizing wave phenomena. *Journal of Science Education and Technology*, *8*(3), 197–209.
Snir, J., Smith, C., & Grosslight, L. (1995). Understanding in the science classroom: Integrating laboratory experiments, students and computer models, and class discussion in learning scientific concepts. In D. N. Perkins, J. L. Schwartz, M. M. West, & M. S. Wiske (Eds.), *Software goes to school: Teaching for understanding with new technologies* (pp. 233–257). Oxford: Oxford University Press.
Tatli, Z., & Ayas, A. (2013). Effects of a virtual chemistry laboratory on students' achievement. *Educational Technology & Society*, *16*(1), 159–170.
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*(4), 423–459.
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*(2), 149–173.
White, B. Y. (1993). Thinkertools: Causal models, conceptual change, and science education. *Cognition and Instruction*, *10*(1), 1–100.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.
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*(1), 25–42.
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*(2), 120–132.
Zacharia, Z. C., & Anderson, O. (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*(4), 425–430.
Zacharia, Z. C., & de Jong, T. (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.
Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, *21*(3), 317–331.
Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching*, *45*(9), 1021–1035.

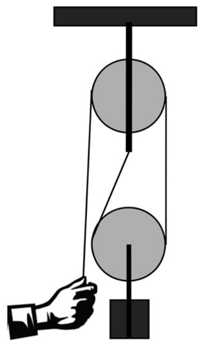
Appendix

Pulley test

Instructions: Choose only *one* letter to indicate your answer for each question.

Important: All of the situations are in an environment with ***no friction***, unless otherwise stated in the question.

(1) Which condition will require *less applied force* to lift an object to a height of 1 metre – using the pulley shown or lifting the object straight up by hand?



A Using the pulley
B Lifting it straight up
C Both using the pulley or lifting it straight up require the same applied force
D Not enough information to decide

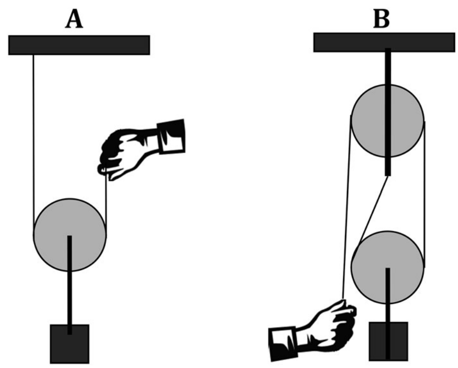
(2) You used a single fixed pulley to lift a watermelon to your tree house. If you use a single movable pulley to move the same watermelon into the same tree house the *applied force* needed would:
A Increase
B Decrease
C Stay the same
D Not enough information to decide

(2a) Explain why.

(3) If friction is a factor, which of the following will require *less applied force* to lift a load to a height of 2 metres using a single fixed pulley?
A A well-oiled pulley
B A rusty pulley that needs to be oiled
C Both a well-oiled pulley and a rusty pulley will require the same applied force
D Not enough information to decide

(3a) Explain why.

(4) Which of the following pulley set-ups will require *less applied force* to lift the load?

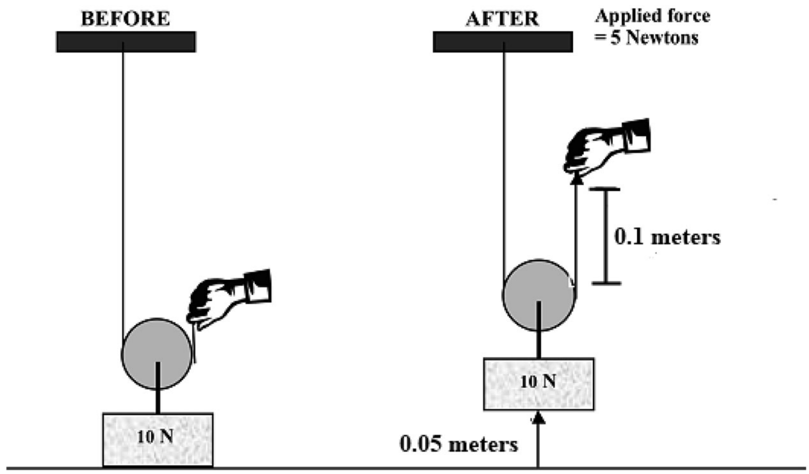


A Pulley set-up A
B Pulley set-up B
C Both A and B will require the same applied force
D Not enough information to decide

(4a) Explain why.

(5) Which one of the following pulley set-ups will require the *least applied force* to lift a load to the same height?
A One fixed pulley
B One movable pulley
C A double compound pulley
D Two fixed pulleys

(6) Below is a picture of a load being lifted with the help of a pulley.



(6a) The distance the load moves is:
A 0.05 metres
B 0.1 metres

(6b) The distance pulled is:
A 0.05 metres
B 0.1 metres

(7) You used a single fixed pulley to lift a watermelon to your tree house. If you use a single movable pulley to move the same watermelon into the same tree house the *distance pulled* would:
A Increase
B Decrease
C Stay the same
D Not enough information to decide

(7a) Explain why.

(8) Describe in your own words the relationship between *applied force* and *distance* in a pulley.

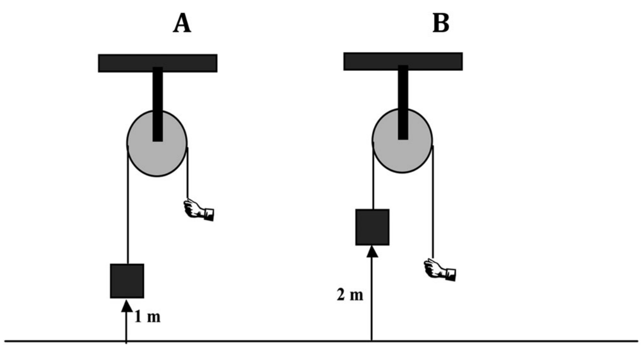
(9) You used a single fixed pulley to lift a watermelon to your tree house. If you use a single movable pulley to move the same watermelon into the same tree house the amount of *work* done would:

A Increase
B Decrease
C Stay the same
D Not enough information to decide

(9a) Explain why.

(10) How does the trade-off between *applied force* and *distance* affect the *work* done when using a pulley?

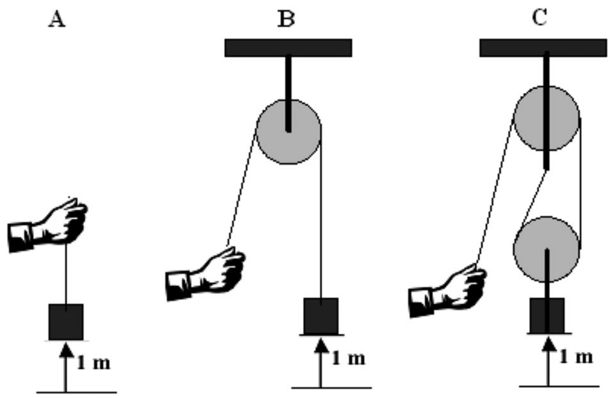
(11) Jacob uses a fixed pulley to lift a box 1 metre. He then uses the same pulley to lift an identical box 2 metres. In which condition did Jacob do *more work?*



A In A when lifting the box 1 metre
B In B when lifting the box 2 metres
C The amount of work done was the same in both A and B
D Not enough information to decide

(11a) Explain why.

(12) Alison lifts a box straight up by hand to a height of 2 metres. Boris uses pulley B to lift same box to a height of 2 metres. Carla uses pulley C to lift the same box to a height of 2 metres. What can you say about the amount of *work* being done?

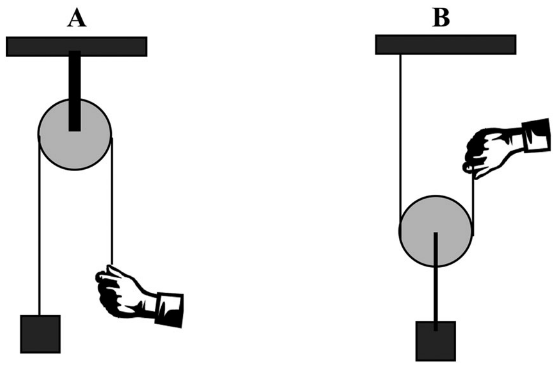


A Alison is doing more work in situation A lifting the box by hand
B Boris is doing more work using pulley B
C Carla is doing more work using pulley C
D Alison, Boris, and Carla are doing the same amount of work

(13) Which one of the following pulley set-ups will give the *most mechanical advantage*?
A One movable pulley
B One fixed pulley
C A double compound pulley
D Two fixed pulleys

(13a) Explain why.

(14) Which one of the following pulley set-ups will give *more mechanical advantage*?



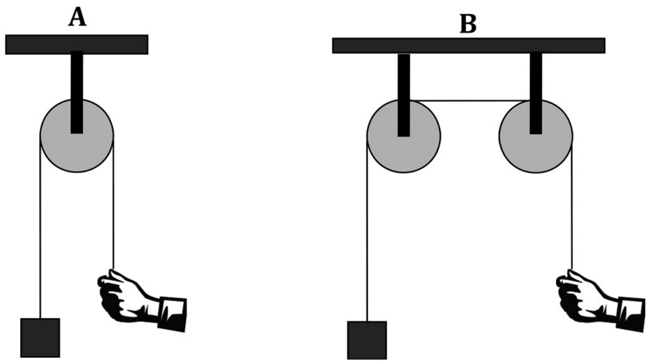
A Pulley set-up A
B Pulley set-up B
C Pulley set-up A and pulley set-up B will give you the same mechanical advantage
D Not enough information to decide

(14a) Explain why.

(15) What is the best way to *increase mechanical advantage* when using a pulley system?

(16) When *mechanical advantage* is increased in a pulley system, what can you say about the *applied force* required to lift an object?

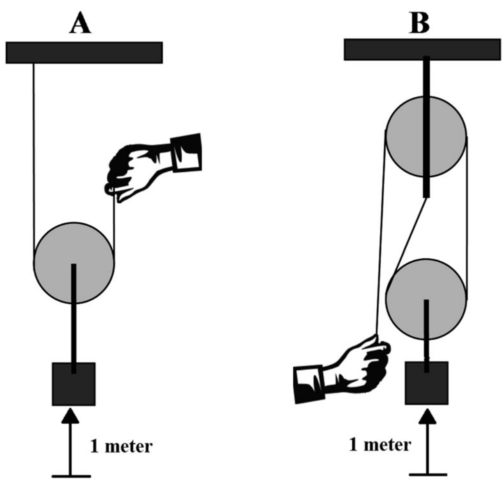
(17) Which one of the following pulley set-ups will give *more mechanical advantage*?



A Pulley set-up A
B Pulley set-up B
C Pulley set-up A and pulley set-up B will give you the same mechanical advantage
D Not enough information to decide

(17a) Explain why.

(18) Louis lifts a box 1 metre using pulley set-up A. Toby lifts an identical box to the same height using pulley set-up B. After the boxes have been lifted, which box has *more potential energy*?



A The box in pulley set-up A
B The box in pulley set-up B
C Both boxes have the same potential energy
D Not enough information to decide

(19) Henry uses a double compound pulley to lift a box to a height of 1 metre. How does the amount of *work* done compare to the *potential energy* of the box after it has been lifted?

A The work done is greater than the potential energy
B The work done is less than the potential energy
C The work done is the same as the potential energy
D Not enough information to decide