

Effects of Simulation-based Formative Assessments on Students' Conceptions in Physics

Mihwa Park ^{1*}

¹ Texas Tech University, USA

Received 14 December 2018 • Revised 26 January 2019 • Accepted 1 February 2019

ABSTRACT

Background: The paper presents effects of simulation-based formative assessments on students' conceptions in physics. In the study, two topics—motion in two dimensions and conservation of energy—were selected to explore students' conceptions in physics, and related assessment tasks incorporating computer simulations and formative assessment questions were developed.

Material and methods: The participant students were first-year college students with majors related to science or engineering. Analytic rubrics were developed to capture the students' normative and non-normative ideas revealed in their responses, and a holistic rubric was applied to categorize the responses into four response models.

Results: The results demonstrated that, overall, students predicted and explained the given scientific phenomena with more valid scientific ideas after experiencing a computer simulation. However, the results also indicated that students' non-normative ideas were still present even after experiencing computer simulations, especially when they were required to consider an abstract scientific concept such as energy dissipation.

Conclusions: The finding can be explained with knowledge-in-piece perspectives (diSessa, 1993), that students' naïve knowledge is fragmented, and thus they do not demonstrate a coherent understanding of abstract science concepts across different situations.

Keywords: computer simulation, misconception, conceptual understanding, formative assessment

INTRODUCTION

Online and blended learning have become increasingly important components in higher education. Many studies have investigated online education outcomes compared to those of traditional classes across a variety of subjects (e.g., Navarro & Schoemaker, 2000; Tallent-Runnels et al., 2006). Nguyen (2015) reviewed the literature, paying particular attention to meta-analyses studies on effects of online learning, and found robust evidence suggesting that online learning was as effective as traditional learning. Especially in science disciplines in higher education, although fully online college-level science courses are not yet commonly offered, the use of online homework has steadily increased over the past two decades (Bonham, Deardorff, & Beichner, 2003; Richards-Babb et al., 2011; Penn, Nedeff, & Gozdzik, 2000), which can be considered as an online learning component to a traditional class. In physics, Cheng et al. (2004) found that students who were assigned online homework showed similar performance results compared to those of students who completed paper-and-pencil homework. Bonham, Deardorff, and Beichner (2003) reported a similar result, that two homework methods—online and traditional paper homework—did not show significant differences on student learning outcomes. Even though previous studies have found that online homework is neither superior nor inferior to traditional homework, use of online homework is becoming a major component of college-level science courses (Bonham, Deardorff, & Beichner, 2003). The increasing popularity of online homework might be due to the fact that there are many online homework systems available; instructors

Contribution of this paper to the literature

- This study contributes to the existing literature on using simulation-based formative assessments to promote students' conceptual learning in science.
- The results showed that simulation-based formative assessments had a positive effect on students' conceptual learning.
- The study, however, also argues that the improvement might be the case only in certain situations. In the study, students demonstrated better abilities to predict and explain the given scientific phenomena after running a simulation, whereas, when an abstract concept was required to apply to answer questions, their abilities regressed.

can easily collect students' homework and sometimes grade it automatically, especially in the case of multiple-choice questions, which gives the instructors more time for class preparation. College-level introductory science courses, especially, are often offered in large classrooms and taught in a traditional lecture-based format (Pascarella & Terenzini, 1991; Springer, Stanne, & Donovan, 1999). Instructors of these courses seek new methods that will keep students engaged in the class and in the learning process, but do not take much time in terms of collecting and grading homework. In these cases, an online homework system might be of interest to instructors as an alternative venue.

There is no doubt that formative assessment is an important component in education. The main role of formative assessment is diagnosing students' learning needs in order to improve their learning outcomes. There are an overwhelming number of studies showing the positive impact of formative assessment on student achievement in science. For example, Black and William (1998) reviewed more than 250 books and articles, and concluded that formative assessment has significant positive effects on student achievement.

Although the effects of formative assessment have been studied extensively, formative assessment as an online complement to the traditional face-to-face class has not been much illuminated. Gikandi, Morrow, and Davis (2011) asserted the importance of using online formative assessment in order to create learner-centered learning environments (Pachler et al., 2010); there was, however, a lack of studies on online formative assessment. Gikandi et al. (2011) reviewed published articles and reports regarding applications of online formative assessment in online and/or blended higher education contexts, and concluded that online formative assessment is beneficial to improve learners' engagement with valuable learning experiences. They also identified formative feedback and embedded authentic assessment activities within the teaching and learning processes as key factors for successful online formative assessment. To be specific, using online tools such as computer-based simulations or collaborative inquiry provides students opportunities to engage in activities and problem-solving relevant to their real-world situations (Crisp & Ward, 2008; Mackey, 2009). Thus, online formative assessments should be designed to encourage and promote student learning experiences through authentic tasks (Gikandi, Morrow, & Davis, 2011).

Computer simulations are programs presenting a representation of an authentic system, a phenomenon, or a process (de Jong & van Joolingen, 1998). Studies found that computer simulations promoted students' engagement in observing and exploring phenomena (Srisawasdi & Kroothkeaw, 2014), which offered students opportunities to promote conceptual change in science (Rutten, van Joolingen, & ven der Veen, 2012; Trundle & Bell, 2010).

In the current study, computer simulations and online formative assessment were incorporated as integral parts of pedagogy for enhancing students' conceptions in science through which they could perform their own investigations and develop conceptual understanding. Although simulation-based learning and formative assessment have been studied widely, there still is a lack in studies focusing on how online formative assessment incorporating computer simulations enhance college students' conceptions in physics. The research hypothesis is that computer simulation-based formative assessments will have a positive impact on students' conceptions in physics. The specific research question for this study is: "To what extent do simulation-based online formative assessments have an influence on students' conceptions in physics?"

Background and Review of Literature

Student knowledge in science

It has been generally agreed that students often have misconceptions in science that differ from expert conceptions (Bransford, Brown, & Cocking, 2000; Hammer, 1996; Treagust, 1988). An effective strategy to change their conceptions to scientific ones is to ask students to make predictions about various situations, then explain the reasons for their predictions. This elicits students' pre-existing understanding; and they can then be provided opportunities to build or challenge their initial understanding (Bransford, Brown, & Cocking, 2000).

Although student misconceptions have been extensively studied during past decades, there has been a debate over students' knowledge structure, as to whether their knowledge is theory-like or fragmented. Theory-like (framework theory) perspectives hypothesize that students' naïve knowledge structures are coherent, and students explain scientific phenomena consistently using the domain-specific structures in everyday life (Carey, 2009; Chi, 2005; Gelman, 1990; Ioannides & Vosniadou, 2002). On the other hand, knowledge-in-pieces perspectives assert that students' knowledge involves "phenomenological primitives" or "p-prims" (diSessa, 1993). This perspective posits that students' naïve knowledge is fragmented, in that their knowledge elements are not organized by overarching theories, but rather their naïve knowledge is a repertoire of multiple quasi-independent knowledge elements (diSessa, 2002). Thus, conceptual change is a process of restructuring and integrating knowledge elements by adding new elements to the existing ones to make appropriate connections among the knowledge elements (diSessa, 2002; Linn, Clack, & Slotta, 2003).

Ozdemir and Clark (2009) addressed the debate over students' knowledge structure using the same physics questions used by diSessa, Gillespie, and Esterly (2004) and Ioannides and Vosniadou (2002). They applied two coding schemes representing theory-like perspectives (Ioannides & Vosniadou, 2002) and knowledge-in-pieces perspectives (diSessa et al., 2004) to investigate if students demonstrated a coherent understanding of force meaning across questions, and found that their results supported knowledge-in-pieces knowledge structure over a framework theory.

Regarding student conceptions in the two scientific topics addressed in the study; projectile motion and conservation of energy, many empirical studies have been done to investigate students' misconceptions. For example, Hynd, Alvermann, and Qian (1997) summarized students' misconception in projectile motion that students believe that an object will move forward until its forward motion overpowers gravity, then begin to deviate downward once a force implanted in the object dissipates. They also do not believe that two different projectiles will reach the ground at the same time in a non-frictional situation. Many people believe that moving objects possess an internal force or impetus that keeps the object in motion; and the moving objects eventually slow down or stop as their impetus gradually dissipates (McClosky, 1983). Also, students believe that heavier objects fall faster; or that gravity finally acts on an object after its impetus is all dissipated (Gunstone & White, 1981). In regard to students' conception of energy conservation, Park and Liu (2019) found that the conservation of energy was significantly more difficult for students to understand than other energy aspects such as energy form or transfer across all science disciplines. Many students just recall the law of conservation of energy and fail to apply the law in solving science problems (Goldring & Osborne, 1994) or explaining living phenomena (Barak, Gorodetsky, Chipman, 1997). Tatar and Oktay (2007) reviewed studies on students' conceptions on energy conservation and summarized that students believe that energy is used up or lost; energy degradation is opposite to energy conservation.

In order to reveal students' misconceptions in science, various methods have been used including interviews, open-ended questionnaires, or ordinary multiple-choice tests. Although interviews are effective for investigating students' misconceptions in science, a large amount of time is required to train interviewers and interview students (Gurel, Eryilmaz, & McDermott, 2015). Treagust (1988) suggested two-tiered test items to overcome those difficulties and effectively diagnose students' misconceptions. A two-tiered item is composed of two related questions: the first tier question is a multiple choice question in which respondents should choose only one option out of multiple options from a list and the second tier question asks respondent to provide their reasoning for the answer given to the first question (Treagust, 1988). Peterson, Tregust and Garnet (1989) support that two-tiered items investigate students' understanding and diagnose their conceptions effectively.

Computer simulations in science teaching

Computer simulations can be used in science teaching to help students observe scientific phenomena that we could not accurately or easily do in real life. Through the meta-analysis of studies on using computer simulations over the last decade, Rutten, van Joolingen, and van der Veen (2012) found positive results on student performance, motivation, and attitude for the classrooms in which simulations were used to replace or enhance traditional instructional practices. Smetana and Bell (2012) reviewed published research on the effectiveness of computer simulation, and found that science teaching that incorporated computer simulations produced, in general, better results than traditional teaching. In particular, they indicated that simulations were effective to enhance student content knowledge and to facilitate their conceptual change when the simulations were used to supplement traditional instruction.

Simply providing educational technologies or software without instructional strategies is not likely to result in positive effects on student learning, especially on conceptual change (Trundle & Bell, 2010). In order to promote student conceptual change using computer simulations, Tao and Gunstone (1999) suggested an instructional strategy that makes a prediction about scientific phenomena in order to confront students with discrepant events that contradict their preconceptions; explains the prediction; runs a simulation to test the prediction; and reconciles

the conflict by reflecting on students' conceptions. In physics, Zacharia and Anderson (2003) used a similar strategy investigating the effectiveness of computer simulations on students' abilities to make predictions and to explain the results of physics experiments and concluded that the simulations facilitated students' conceptual change. There are many empirical studies supporting the use of computer simulations in science classrooms in order to effectively facilitate students' conceptual understanding of scientific ideas (e.g., Dori & Barak, 2001; Geban, Askar, & Ozkan, 1992; Huppert, Lomask, & Lazarowitz, 2002; Winn et al., 2006).

Although many studies have focused on the effectiveness of computer simulations employing a pre- and post-test design, fewer studies have been done using performance data collected while students were interacting with simulations (de Klerk, Veldkamp, & Eggen, 2015). In order to collect student performance data, it is necessary to embed formative assessment approaches into computer simulations used in science classrooms. Recently, some studies have been done focusing on performance data of simulation-based assessments. For example, Quellmalz et al. (2012) developed simulation-based science assessments for two middle school science topics, and collected data that included student responses to the assessments and classroom observations. Their findings showed that students performed better on simulation-based assessments than on conventional static assessments. Srisawasdi and Panjaburee (2014) investigated the effect of simulation-based inquiry integrated with formative assessment on students' conceptual learning of buoyancy-driven phenomena. They reported a positive effect of computer simulations embedded in assessments: students who had experienced the simulation-based formative assessments showed significantly higher performance scores than students who had not experienced the assessments. Park et al. (2017) developed simulation-based formative assessments to promote student understanding in high school chemistry, and found a significant effect of the frequency of using computer simulations as formative assessments on students' understanding of the nature of models (Treagust, Chittleborough, & Mamiala, 2002).

Instructors of physics were among the first in the various science disciplines to explore computer simulations in teaching and learning. As a result, there are currently many computer simulations available for various physics topics (Jimoyiannis & Komis, 2001). In this project, computer simulations were incorporated as an integral component in formative assessment to enhance student conceptions in physics, leading to the development of simulation-based formative assessments. The simulation-based formative assessments target high school physics classes and introductory-level college physics courses to assist instructors in assessing student conceptions as well as enhance students' deeper learning of core scientific ideas.

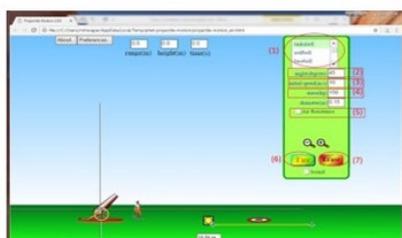
METHODS

Participants

An invitation email to participate in the study was sent to college students who were taking an introductory physics course at a research university in the United States. Initially, 75 students responded to the email. Among those students, 70 students participated in the first task, and 67 students participated in the second task (3 students who participated in the first task did not participate in the second task). The participants were first-year college students whose majors related to science or engineering (e.g., mechanical engineering, chemical engineering, physics, etc.). The physics course syllabus was obtained at the beginning of the semester, and each task was administered right after a relevant topic was taught. The participants were given five days to complete each task online.

Developing Simulation-Based Formative Assessments

The basic structure of the formative assessments consisted of a computer simulation and a series of two-tiered questions and constructed response (CR) questions. The questions were posed while students were running a computer simulation as the system directed. A two-tiered question consists of a simple multiple-choice question as its first tier and a justification question asking students to write their justification for their choice as its second tier (see [Figure 1](#)). Two-tiered questions have advantages over simple multiple-choice questions in that they provide information about students' reasoning or conceptions behind their selected responses (Gurel, Eryilmaz, & McDermott, 2015).



Note: (1)~(5) will be used while you are working on this task.

When you click the “Fire” (6) button, the cannon will shoot an object as you set the variables (angle, speed and mass). If you click the “Erase”, the simulation will be reset.

Q. When air resistance is negligible, a cannon shoots a tank shell and a baseball. Before you run the simulation, speculate which one will go farther. The angle and initial speed are the same for both cases.

- a. Tank shell b. Baseball c. Same

Q. Explain your choice without using a formula.

*Run the simulation (set the angle at 45, initial speed at 10 m/s).

Q. why do you think it happened?

Figure 1. Example questions in Task 1

The formative assessments were designed for online administration that allowed students to use the system at their convenience. A website was created to integrate the two components— computer simulations and formative assessment questions—so that students were able to answer the questions while they were running a simulation on the same web interface.

According to the course syllabus, an object’s motion with the law of motion and mechanical energy with the concept of energy conservation were core ideas to teach in the first half of the semester. In the current study, two topics: (1) motion in two dimensions (projectile motion) and (2) conservation of energy were selected to align with the course syllabus. Computer simulations related to the two topics were selected from the Physics Education Technology (PhET) project, which develops research-based simulations (Perkins et al., 2006) and allows users to embed their simulations on their own websites. PhET simulations were designed to emphasize the connection between physics and everyday life and to facilitate the development of a robust conceptual understanding of physics (Perkins et al., 2006). The selected simulations were (1) Projectile Motion and (2) Energy Skate Park (retrieved on February 2017 from <https://phet.colorado.edu>). After selecting simulations, CR and two-tiered questions related to the simulations were developed. When developing questions, conceptual understanding in physics were targeted for assessment, thus students were not asked to calculate any values or to demonstrate their mathematical competence. The questions first asked students to predict what would happen in a given situation; then the assessment system asked students to run a simulation and posed questions asking them to explain the observed phenomena (Figure 1). In the process, the website system ensured that simulations were not enabled before students responded to the questions asking them to make a prediction. Each task included a series of questions asking students to predict and explain scientific phenomena. To be more specific, after students completed the first set of questions—making predictions and explaining results before and after running a simulation—they were asked to answer a new set of questions. The new set of questions used similar activities, but different simulation settings. For example, the first simulation in Task 1 was a cannon shooting a tank shell and a baseball. After students answered a series of questions related to the first simulation, they were asked to answer a second set of questions, which were associated with a simulation shooting a piano instead of a tank shell. After answering these questions, students interacted with more simulations, which were set in different situations such as different initial speeds and initial angles, and to responded questions in order to examine if they could transfer their conceptions to different situations. Task 1 for the topic (1) included 14 questions, and Task 2 for the topic (2) included 17 questions. Two physics professors (one a retired physics professor who had taught introductory physics courses over 20 years, and the other currently teaching an introductory physics course) were asked beforehand to review the content of the questions to ensure their scientific accuracy and relevance to college freshman. The experts conducted their reviews independently. They made specific suggestions on wording to improve a few questions’ clarity, and revisions were made accordingly.

Table 1. Analytic Rubrics for Task 1 (a) and Task 2 (b)*a. Task 1 Scoring Rubric*

Idea Types	Ideas	Idea Descriptions
Normative idea	Effect of initial factors	<ul style="list-style-type: none"> • Mass is not relevant to projectile motion • Initial velocity and/or angle affect(s) projectile motion
	Effect of external factors	<ul style="list-style-type: none"> • Gravity is the only force acting on an object in non-frictional situations • Object's acceleration is constant in non-frictional situations • Air resistance affects an object's projectile motion
Non-normative idea	Effect of initial factors	<ul style="list-style-type: none"> • Mass affects an object's projectile motion • Initial velocity or angle is not relevant to an object's projectile motion • Higher velocity or angle results in a greater acceleration or greater force acting on the object • Object size affects its motion in non-frictional situations
	Effect of external factors	<ul style="list-style-type: none"> • Different amounts of gravity force result in different acceleration rates • A heavier object will have greater air resistance
Off-task	Off topic, repetition	<ul style="list-style-type: none"> • Repeats the question • Off topic

b. Task 2 Scoring Rubric

Idea Types	Ideas	Idea Descriptions
Normative ideas	Effect of initial factors	<ul style="list-style-type: none"> • Mass is not relevant to its speed/peak height in non-frictional situations
	Effect of external factors	<ul style="list-style-type: none"> • Gravity is the only force acting on an object in non-frictional situations • Object's acceleration is constant in non-frictional situations • Friction affects an object's motion
	The concept of energy	<ul style="list-style-type: none"> • Mechanical energy is conserved in non-frictional situations (the law of energy conservation) • Mass is related to the amount of energy • Some energy is dissipated as thermal energy (heat)
Non-normative ideas	Effect of initial factors	<ul style="list-style-type: none"> • Mass affects an object's motion
	Effect of external factors	<ul style="list-style-type: none"> • Different amounts of gravity force result in different acceleration rates • A heavier object will have greater air resistance • Friction doesn't affect an object's motion
	The concept of energy	<ul style="list-style-type: none"> • The greater amount of energy results in greater speed, and vice versa • Mass is not related to the amount of energy • When mass changes, either PE or KE will change, not both • Heat is an addition to the total amount of energy (the amount of total energy increases when heat is created) • Gravitational potential energy is greater as closer to the ground • Mechanical energy is always conserved in frictional situations (No energy dissipation) • As mass increases, more energy is required to initiate its motion
Off-task	Off topic, repetition	<ul style="list-style-type: none"> • Repeats the question • Off topic

Scoring Rubric

The formative assessments included both two-tiered questions and CR questions; thus, rubrics for scoring students' written responses were necessary. In the study, analytic rubrics were developed to capture students' normative ideas (relevant scientific ideas) and non-normative ideas (misconceptions) revealed in their responses (Table 1). When developing analytic rubrics, student responses were first analyzed qualitatively using open-coding, as codes were developed to explore themes in student responses as they emerged. After a line-by-line analysis of each written response, response codes were categorized into *Ideas* as they shared commonalities, such as focusing on an object's initial factors (mass, shape, velocity, etc.), external factors (external force, air resistance, etc.), or the concept of energy (energy form, transformation, dissipation, and conservation). Those ideas were then classified into normative and non-normative ideas (*Idea types*). In the rubrics (Table 1), response codes were presented as *Idea descriptions*. Some students wrote off-topic, nonrelevant responses such as "I don't know," "I saw it in the simulation," or "I learned it from [the] last lecture." In those cases, student responses were classified into *Off-task*. The analytic rubric differs from the holistic rubric in that it breaks the holistic evaluation of students' explanations into normative scientific ideas or non-normative ideas assessed by questions. After developing scoring rubrics, each written response was quantified by tabulating the frequency of normative and non-normative ideas. During the process, two raters independently analyzed student responses and demonstrated high inter-rater

Table 2. Holistic Rubric

Model	Score	Description
Model 1	0	Off-task
Model 2	1	Non-normative ideas only
Model 3	2	Co-existence of normative and non-normative ideas
Model 4	3	Normative ideas only

reliability (kappa coefficients > 0.8). In cases of disagreement between raters, the discrepancies were discussed collectively until agreement was reached. Multiple-choice questions were scored dichotomously (i.e., “1” for correct answers and “0” for incorrect ones).

After analyzing students’ responses to questions with the analytic rubric, a holistic rubric was applied to categorize students’ responses into four different response models (Table 2). This method is similar to that of Moharreri, Ha, and Nehm (2014), who developed a web portal, EvoGrader, to grade students’ written explanations in evolution. EvoGrader reports student holistic reasoning models: pure non-normative ideas (non-normative ideas only), mixed ideas (non-normative ideas and normative ideas), and pure scientific ideas (normative ideas only). In the current study, the holistic rubric includes the off-task response model (Model 1) along with the three response models (Moharreri et al., 2014).

Data Analysis

After scoring student responses, the scored data were subjected to analysis using the Rasch model. In particular, the partial credit Rasch model (Masters, 1982) was applied to incorporate differing numbers of response opportunities for different items (questions) in a test, including dichotomously scored items and partial credit scored items (Bond & Fox, 2007). As a result of applying the partial credit Rasch model, an item difficulty estimate for each item was produced in a logit scale. The larger the item difficulty estimate, the more difficult for students to achieve the maximum score on that item. Item difficulty estimates produced by the Rasch model are linear measures on the interval scale, allowing for direct comparisons between item difficulties. Thus, item difficulty estimates from the partial credit Rasch model were used to compare student performances on each question both before and after running a simulation. Winsteps 3.81 (Linacre, 2014) was used for the Rasch model analysis.

RESULTS

The reliability coefficients indicated by Cronbach’s alpha was 0.70 for both tasks.

Partial Credit Rasch Modeling Analysis

The Rasch model provides fit statistics presenting how well each item fits within the underlying construct (Bond & Fox, 2007) as evidence of construct validity of measures. Item fit statistics include the mean square residual (MNSQ) and the standardized mean square residual (ZSTD). Rasch modeling analysis also produces INFIT statistics, giving more weight to better fitting responses, and OUTFIT statistics, giving no weight over all responses for MNSQ and ZSTD. Using a commonly used criterion for acceptable fit of items, MNSQ within the range of 0.7 to 1.3 and ZSTD within the range of -2.0 to +2.0, one item (Q1) in Task 1 was found with two statistics out of the acceptable ranges. Thus, Q1 was excluded from the analysis. Question 1 was a multiple-choice question asking, “Between a tank shell and a baseball, which one will go farther when they are shot under the same conditions? Even though it was removed, students’ written responses to explain their predictions (Q2) were included in the analysis. After the deletion of the question, 13 questions were subjected to the Rasch modeling analysis to produce item difficulties. In case of Task 2, Q5 deviated from the item response function, but the other questions fit the Rasch model well. Q5 asked students if they observed how the skater’s highest speed changed as her mass increased. All of the students chose the same option: that the skater’s highest speed stayed the same. No one selected the other options, which were that the highest speed increased or decreased. If a question did not contribute to differentiate students’ ability at all, the Winsteps software excluded the question in the analysis. Thus, Q5 was automatically dropped from the analysis.

Effects of Simulation-Based Formative Assessments

Each task was designed for the students to make a prediction about what would happen in a given situation, then explain the prediction. After explaining their prediction, they ran a simulation to test it, and wrote an explanation of the scientific phenomena. After they completed those questions, the students were asked to predict

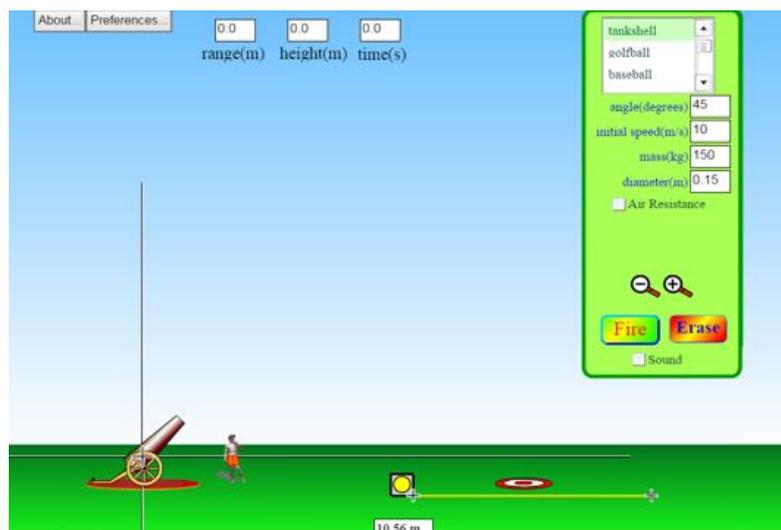


Figure 2. Projectile motion simulation. Source: <https://phet.colorado.edu/en/simulation/projectile-motion>

Table 3. Descriptions of Formative Assessment Questions in Task 1

Descriptions of questions and simulations	Question number	Question format	Before/After ¹	Item difficulty (logit)
(shooting a tank shell and a baseball without friction in the same condition of initial speed and angle) When a cannon shoots a tank shell and a baseball, which one goes farther?	2	SW ²	Before	1.19
(shooting a piano and a baseball without friction in the same condition of initial speed and angle) When a cannon shoots a piano and a baseball, which one goes higher?	3	SW	After	0.44
(shooting two tank shells without friction and with the same initial angle) When a cannon shoots two tank shells in a row, which one goes farther? The first tank shell's initial speed is two times greater than the second one.	4	MC ³	Before	-0.12
	5	SW	Before	0.06
	6	SW	After	0.05
(shooting two tank shells without friction and with the same initial speed) When a cannon shoots two tank shells in a row, which one goes farther? The first tank shell's initial angle is 45 and the second one's initial angle is 10.	7	MC	Before	-1.65
	8	SW	Before	-0.11
	9	SW	After	-0.01
Describe what factors can be changed and how they can be changed to increase the horizontal range of the projectile object.	10	MC	Before	-0.25
	11	SW	Before	0.08
Elaborate on your explanations for Q12 using the idea of the object's velocity and acceleration in the vertical and horizontal directions	12	SW	After	0.11
Elaborate on your explanations for Q12 using the idea of the object's velocity and acceleration in the vertical and horizontal directions	13	SW	After	-0.20
Elaborate on your explanations for Q12 using the idea of the object's velocity and acceleration in the vertical and horizontal directions	14	SW	After	0.42

Note. 1. Before/After; Before—a question was prompted before a running a simulation, After—a question was prompted after a running a simulation

2. SW: Short written response question

3. MC: Multiple-choice question

and explain a different situation to assess if they were able to transfer their conceptions to a new situation, and/or to generalize what they had learned, using ideas that they had experienced in the simulation.

Task 1. Task 1 included a simulation (Projectile Motion in PhET, Figure 2) providing a situation where a cannon shoots an object. The simulation allows users to choose objects (with different shapes and mass) to shoot, and to set an initial angle and speed.

Table 3 presents descriptions of questions and simulation settings, and shows item format and item difficulty estimates before and after experiencing a simulation. At the end of the task, students were asked to generalize what factors should be changed to increase the horizontal range of the projectile object (Q13); this was in order to examine if the students could integrate what they had learned through the activities. Q14 provided another opportunity for students to explain, using the vector concept in velocity and acceleration. This was to see if the students could apply the concept of vector in motion, which was important to understanding the projectile motion.

Three multiple-choice questions demonstrated lower item difficulties as compared to the short written response questions (e.g., second-tier questions or CR questions). The findings indicated that students were not always able to give an explanation using normative scientific ideas, even when they had selected the correct option in the

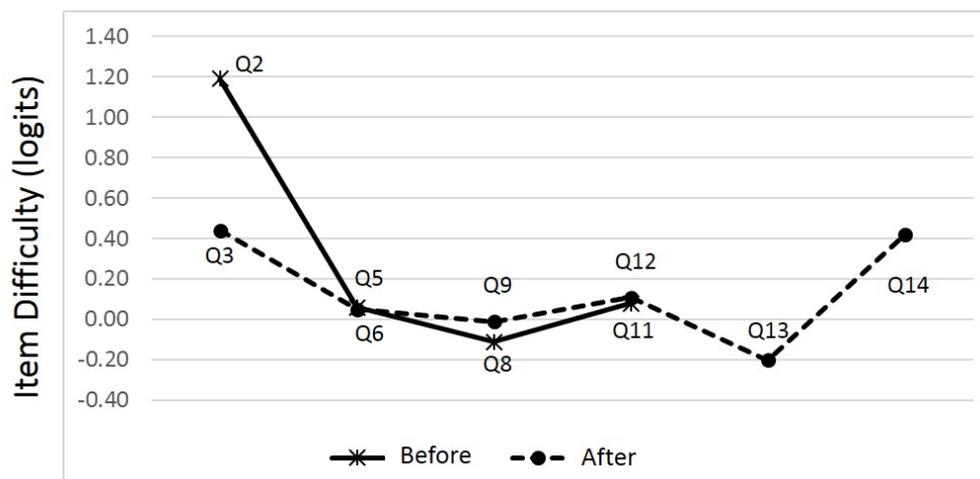


Figure 3. Task 1 item difficulties before and after running a simulation

multiple-choice question. For example, in the cases of Q10 and Q11, some students who chose the multiple-choice option that the first tank shell would go farther (Q10) wrote that gravity would more directly affect the tank shell with the angle of 10 degrees, or that, because its angle was greater, it would result in more area to cover. As such, in order to compare the students' responses before and after running the simulations, it was necessary to compare the students' responses to the same format of questions. Thus, item difficulties of the short written response questions were compared before and after a running simulation (Figure 3).

As seen in Figure 3, Q3's item difficulty was lower than Q2's, indicating that the students had less difficulty answering the question after running the simulation. This supported the effectiveness of computer simulations that helped students explain scientific phenomena using normative scientific ideas. Except for Q2 and Q3, the findings showed that item difficulties of questions posed after running a simulation—asking the students to explain observed scientific phenomena from the simulation (explanation questions)—were not always lower than those of questions students answered before a running the simulation—asking them to make a prediction about a given situation (prediction questions). Specifically, the explanation questions showed similar difficulty locations (Q6, Q9, and Q12) to their paired prediction questions (Q5, Q8, and Q11, respectively). When comparing Q2 and Q5 (prediction questions), although they were designed around a similar situation, Q5's item difficulty was less than Q2's item difficulty (the difference between the two item difficulties was 1.13 logits). The differences between the two questions were that the cannon would shoot a piano (Q5) instead of a tank shell (Q2); additionally, Q5 asked which object would go higher and Q2 asked which one would go farther. The result showed that after experiencing the first set of questions with a simulation, students had less difficulty making a correct prediction for the given situation. Students had seen in the simulation that two different objects (with different mass and shapes) would travel exactly the same, and revised or confirmed their initial thoughts in Q3 (explanation question), which helped them make a correct prediction in answer to the question. In the case of two objects with either different initial velocities or different initial angles, the item difficulties for both the prediction and explanation questions were located similarly. Their locations were also similar to those of the previous question set (Q5 and Q6), implying that students might demonstrate their conceptions consistently across slightly different situations asking about the same scientific ideas (factors affecting an object's projectile motion). Q13, which asked students to generalize their conceptions on projectile motion, showed the lowest difficulty location among all of the short written response questions, which might imply that, after running all of the simulations, their explanations were more scientifically valid. Q14's difficulty location explained that the students had more difficulty explaining projectile motions using velocity and acceleration ideas in two different directions, which required them to apply a new concept, vector.

Table 4 presents descriptive statistics for the short written response questions in Task 1 using the holistic rubric.

Table 4. Descriptive Statistics for Task 1 Short Written Response Questions

Question	Frequency (%)				Mean Score
	Model 1 Score 0	Model 2 Score 1	Model 3 Score 2	Model 4 Score 3	
Q2	0 (0)	24 (34.3)	0 (0)	46 (65.7)	2.31
Q3	6 (8.6)	5 (7.1)	1 (1.4)	58 (82.9)	2.59
Q5	2 (2.9)	11 (15.7)	0 (0)	57 (81.4)	2.60
Q6	3 (4.3)	2 (2.9)	0 (0)	65 (92.9)	2.81
Q8	2 (2.9)	2 (2.9)	1 (1.4)	65 (92.9)	2.84
Q9	2 (2.9)	5 (7.1)	4 (5.7)	59 (84.3)	2.71
Q11	2 (2.9)	10 (14.3)	3 (4.3)	55 (78.6)	2.59
Q12	2 (2.9)	13 (18.6)	1 (1.4)	54 (77.1)	2.53
Q13	1 (1.4)	2 (2.9)	15 (21.4)	52 (74.3)	2.69
Q14	5 (7.1)	7 (10.0)	3 (4.3)	55 (78.6)	2.54

In the case of Q3, the frequency of Model 4 (normative ideas only) responses increased and the frequency of Model 2 (non-normative ideas only) responses decreased, compared to Q2. Q5 and Q6 showed similar patterns: that the number of students' non-normative responses decreased, and their normative responses increased, after running a simulation. The result demonstrated that more students' responses included scientifically normative ideas instead of non-normative ones after experiencing a computer simulation. Q8 and Q11 asked students to make predictions about the projectile motions of the same object, but with different initial velocities or angles. In general, the students' predictions were scientifically correct (Model 4 responses) when the objects' initial velocities were different but other factors were the same (Q8), while the number of correct responses decreased in a situation that changed the initial angles (Q11). As Table 4 presented, the number of the students' non-normative ideas increased in Q11, compared to Q8. Even after a simulation, the number of Model 2 responses did not decrease (Q12). Interestingly, in Q9, the number of responses in Models 2 and 3 (non-normative ideas only and mixed ideas, respectively) increased, compared to Q8 responses. Those results did not support the effectiveness of using simulations for student scientific learning. Examples of responses that changed from Model 4 to Models 2 or 3, between Q8 and Q9, and between Q11 and Q12 are presented below.

Student A's responses

Q8: x-displacement formula depends on the initial speed of the object. Tank shell's A initial speed was greater therefore will go higher. (Model 4)

Q9: The greater the speed the greater the acceleration and the object will go farther. (Model 2)

Student B's responses

Q8: The faster the velocity is while having the same angle will make the object go further in the x direction. (Model 4)

Q9: The slower object did not have enough speed to combat gravity the way an object at a faster speed would. (Model 3)

Student C's responses

Q11: They are fired at the same speed and 45 degrees is the optimal angle for an object to travel the most distance. The higher angle will travel higher and allow for more time in flight. (Model 4)

Q12: It (lower angle one) spent less time accelerating up. (Model 2)

These examples illustrate that the students used non-normative ideas to given an explanation after running a simulation. They were allowed to differentiate their responses from previous ones, which gave students an opportunity to express their non-normative ideas, implying that they already had the non-normative ideas before running the simulation. It is also possible that the simulation helped the students to uncover their non-normative ideas.

With the exceptions of those responses, the findings support the idea that computer simulations, coupled with a series of formative assessment questions, were effective for students to find their non-normative ideas and revise their answers to make them scientifically valid. After running a couple of similar simulations, the number of student response models did not show much difference between prediction questions and explanation questions.

Task 2. The Task 2 simulation (Energy Skate Park in PhET, Figure 4) was designed to demonstrate the concept of the conservation of mechanical energy using kinetic energy (KE) and gravitational potential energy (PE). The simulation allows users to change a skater's mass and track shapes and to include or remove friction.



Figure 4. Energy skate park simulation. Source: <https://phet.colorado.edu/en/simulation/energy-skate-park-basics>

Table 5. Descriptions of Formative Assessment Questions in Task 2

Descriptions of questions and simulations	Question number	Question format	Before/After ¹	Item difficulty (logit)
(U-Shaped track without friction) What will happen to the skater's highest speed if you increase her mass?	1	MC ²	Before	2.19
	2	SW ³	Before	0.39
(U-Shaped track without friction) How will skater's total mechanical energy change when her mass increases?	3	MC	Before	-0.06
	4	SW	Before	0.63
(U-Shaped track without friction) How did the skater's total mechanical energy change according to the bar graph in the simulation?	6	MC	After	-0.06
(U-Shaped track without friction) Explain why her speed changed or didn't change.	7	SW	After	-0.15
(Ramp-shaped track without friction) How will the skater's speed be different when she arrives at the ground as her mass increases?	8	MC	Before	-0.44
	9	SW	Before	0.14
(U-Shaped track with friction) How will the skater's peak height change as she is skating along the track?	10	MC	Before	-2.18
	11	SW	Before	-0.88
(U-Shaped track with friction) How will the skater's total mechanical energy change as she is skating along the track?	12	MC	Before	0.65
	13	SW	Before	0.70
(U-Shaped track with friction) Explain the relationship between the amount of thermal energy changes, the skater's peak height, and total amount of energy.	14	SW	After	-0.46
(Ramp-shaped track with friction) Speculate the amount of kinetic energy of the skater when she arrived at the ground.	15	MC	Before	0.11
	16	SW	Before	0.11
(Ramp-shaped track with friction) Speculate about how the amount of kinetic energy of the skater will change as she travels along the ground after leaving the track.	17	SW	Before	-0.69

Note. 1. Before/After; Before—a question was prompted before a running a simulation, After—a question was prompted after a running a simulation

2. MC: Multiple-choice question

3. SW: Short written response question

Table 5 presents detailed descriptions of the questions and item difficulty estimates before and after experiencing a simulation. At the end of the task, students were asked to speculate how the amount of kinetic energy would change as the skater traveled along the ground after leaving the track. The question was intended to see if students would be able to transfer their energy conceptions to flat ground.

Task 2 questions included more prediction questions than explanation questions. As seen in Task 1, the explanation questions' difficulties were similar to the prediction questions' difficulties after students experienced simulations; however, student responses seemed to be affected when they were required to consider a new concept for the given situation. In that regard, Task 2 involved a greater variety of situations requiring students to consider new concepts than Task 1. Although students were not asked to explain observed phenomena every time, the website ensured that after answering prediction questions, students should run the simulation associated with the prediction question in order to move to the next set of questions. **Table 5** showed that Q7 and Q14 (explanation questions) were easier for the students than their paired prediction questions (Q2 and Q13, respectively).

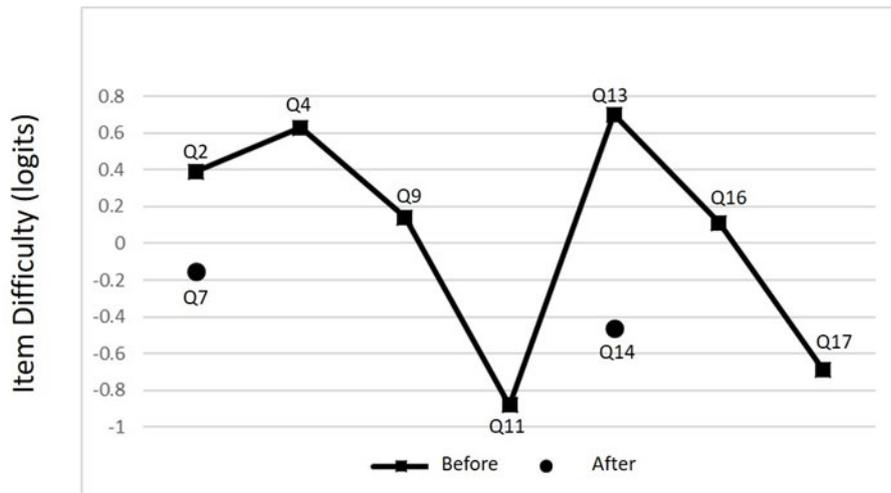


Figure 5. Task 2 item difficulties of questions before and after running a simulation

Table 6. Descriptive Statistics for Task 2 Short Written Response Questions

Question	Frequency (%)				Mean Score
	Model 1 Score 0	Model 2 Score 1	Model 3 Score 2	Model 4 Score 3	
Q2	2 (3.0)	24 (35.8)	27 (40.3)	14 (20.9)	1.79
Q4	0 (0.0)	18 (26.9)	11 (16.4)	38 (56.7)	2.30
Q7	3 (4.5)	13 (19.4)	6 (9.0)	45 (67.2)	2.39
Q9	0 (0.0)	11 (16.4)	8 (11.9)	48 (71.6)	2.55
Q11	1 (1.5)	3 (4.5)	3 (4.5)	60 (89.6)	2.82
Q13	0 (0.0)	20 (29.9)	9 (13.4)	38 (56.7)	2.27
Q14	2 (3.0)	2 (3.0)	14 (20.9)	49 (73.1)	2.64
Q16	3 (4.5)	23 (34.3)	8 (11.9)	33 (49.3)	2.06
Q17	1 (1.5)	9 (13.4)	5 (7.5)	52 (77.6)	2.61

The multiple-choice questions' item difficulties were lower than or similar to those of their paired short written response questions, with one exception. Q1 asked students to predict how a skater's highest speed changes when her mass increases; only 20 students out of 67 chose the correct option. This is a common misconception for students: that heavier objects fall faster than lighter objects. The findings showed that Q1 was the most difficult question in Task 2 for the participant students.

To investigate the student responses in more depth, the item difficulties and descriptive statistics for the short written response questions are presented in Figure 5 and Table 6.

Students answered Questions 2 and 4 before running the first simulation. After running the first simulation, the explanation question (Q7) showed a lower item difficulty, which supported the idea that the simulation was effective for students in explaining the given scientific phenomena. When students were asked question similar to Q2 but in a different situation (Q9; skating on a ramp-shaped track rather than a U-shaped one), Q9's difficulty decreased. The situation in Q11 was close to what students had actually observed and experienced in real life, asking how a skater's peak height would change on a U-shaped track that had friction. The question demonstrated the lowest item difficulty in the task. However, when students were asked about total mechanical energy changes that they couldn't observe directly in real life, the item difficulty increased (Q13). This question assesses students' understanding of the energy dissipation concept, meaning that in order to answer the question, students should take into consideration that mechanical energy transforms into thermal energy. After they experienced a simulation, the students had less difficulty answering its paired explanation question (Q14), which again supported the idea that the simulation helped students better understand the concept of energy dissipation. Q16 and Q17 were situated on a ramp-shaped track with friction, asking how the skater's KE would change when she arrived at the ground and was traveling on the ground. The two questions' difficulties were lower compared to the previous questions (Q4 and Q13) asking about a similar concept, the skater's mechanical energy. In sum, the findings supported the positive effects of simulation-based formative assessment, not only on explaining given scientific phenomena, but also on students' predictions.

Table 6 presents descriptive statistics for the short written response questions in Task 2, using the holistic rubric (Table 6).

In the case of Q2, more than 75% of the students' responses were categorized as Model 2 or Model 3 responses, indicating that non-normative ideas were prevalent in the participant students. Once students experienced the first simulation, the frequency of Model 4 responses increased, while the frequency of Model 2 and Model 3 responses decreased, meaning that more students became able to explain the observed phenomena using normative ideas. Q13 also elicited many students' non-normative ideas (e.g., the total amount of mechanical energy will be always conserved). After experiencing the simulation, their explanations were improved by including normative ideas (Q14). Noticeably, Model 3 responses increased in Q14 responses compared to Q13 responses, implying that some of responses shifted from the only non-normative response model (Model 2) to the mixed model (Model 3) when the students explained the observed phenomena. Although the frequency of Model 2 responses decreased after running the simulation, as we have seen with Q14, their non-normative ideas emerged again in Q16. Q13 and Q14 were situated on a U-shaped track, while Q16 was in a ramp-shaped track, and they both were designed to examine if students could apply the energy dissipation concept along with the mechanical energy concept. Contrarily, Q17 asked students if they could provide the amount of KE's change on the ground with friction. In this case, more than half of the students' responses were categorized as Model 4 and only nine students' responses were found to be Model 2 responses. The findings indicated that the students had difficulty in applying the energy dissipation concept across different situations, especially when they were required to take gravitational PE into account; however, in a more common situation, such as skating on the ground (no need to consider gravitational PE), more students could explain the given phenomena using valid ideas such as the energy dissipation concept. Below are examples of a student's responses to Q13, Q14, Q16, and Q17, which changed from Model 2 to Model 4 after running a simulation but returned to Model 2 in a different situation (a ramp-shaped track), then demonstrated a Model 4 response in a more familiar situation (the ground).

Student D's responses

Q13: Stay the same because the mechanical energy will increase if the speed increases and the speed stayed the same. (Model 2)

Q14: The thermal energy is the energy lost to friction, which means the mechanical energy has decreased. (Model 4)

Q16: Same, because the energy in the potential gets transferred into kinetic. (Model 2)

Q17: As the skater leaves the track, it will turn into thermal energy which would decrease the kinetic energy. (Model 4)

Overall, the result supports the positive effects of simulations coupled with a series of formative assessment questions for helping students to predict and explain scientific phenomena. However, when students needed to apply an abstract scientific concept, such as energy dissipation, the result was not always promising.

DISCUSSION

As simulations-based assessments become ever more popular in science classes (de Klerk, Veldkamp, & Eggen, 2015), there is a growing need to study how students learn while interacting with the assessments. This study explored the effects of simulation-based formative assessments on students' conceptions in physics, especially focusing on the changes in student responses across different situations. Formative assessment questions were designed to first elicit students' prior knowledge and thinking by asking a prediction question, then to enable them to run a simulation, and then to provide them an opportunity to evaluate their previous answer by explaining the observed phenomena. Bransford, Brown, and Cocking (2000) and Tao and Gunstone (1999) emphasized the elicitation of students' pre-existing understanding and the provision of opportunities for challenging their initial understanding for concept change and conceptual learning in science. After giving students two assessment tasks, an exploratory study was conducted to gain insight into how the students' responses changed in quality while they were interacting with the tasks.

The results showed that, overall, students better explained the given scientific phenomena after they had experienced simulations. Also, after experiencing the first simulation, students were able to make a prediction for a similar situation using valid scientific ideas, which supports the effectiveness of using simulations coupled with formative assessment questions in learning science. This might be due to the possibility that students could add a new knowledge element or refine their existing ones through interacting with the simulations and answering questions. The findings indicated that the simulation-based formative assessment had a positive effect on enhancing students' conceptual development of scientific ideas, which is consistent with previous studies on the effectiveness of simulation-based formative assessment on student conceptual change (Quellmalz et al., 2012; Srisawasdi & Panjaburee, 2015).

However, the findings also indicated that students' non-normative ideas emerged even after experiencing simulations, especially when they were required to consider an abstract concept (e.g., energy dissipation). Previous

studies have found that students have difficulty understanding energy dissipation (e.g., Black & Solomon, 1983) and the conservation of energy (e.g., Boyes & Stanisstreet, 1990; Driver & Warrington, 1985). Duit (2012) mentioned that students are able to understand energy conservation only if energy dissipation is also considered. Recently, research has suggested that energy dissipation and conservation should be thought of in conjunction (Park & Liu, 2016). The findings demonstrated that many students were able to predict and explain scientific phenomena with normative ideas related to energy conservation in non-frictional situations after they experienced the first simulation; however, their non-normative ideas emerged again in frictional situations, especially when they were required to consider both PE and KE. This might be due to the possibility that students predicted and explained scientific phenomena based on their rote memorization of the energy conservation law, e.g., the total amount of PE and KE is always conserved. This finding is consistent with the previous studies above that found that students lack a true understanding of energy conservation, which requires an understanding of energy dissipation.

The current study supports the claim that simulation-based formative assessments are effective to enhance students' learning in science (e.g., Quellmalz et al., 2012; Srisawasdi & Panjaburee, 2015). The study, however, also argues that the improvement might be the case only in certain situations. In the study, students demonstrated better abilities to predict and explain the given scientific phenomena after running a simulation, whereas, when an abstract concept was required to apply to answer questions, their abilities regressed. The finding can be explained with knowledge-in-piece perspectives (diSessa, 1993), that students' naïve knowledge is fragmented, and thus they do not demonstrate a coherent understanding of abstract science concepts across different situations. To predict and explain scientific phenomena coherently, a student must have integrated and organized the knowledge elements (Ozdemir & Clark, 2009).

Overall, although the students produced explanations with normative ideas in certain situations, the current study underlines that the use of computer simulations coupled with guiding questions did not always reduce invalid explanations and non-normative ideas, especially when considering abstract concepts across different situations. In a limited sense, simulation-based formative assessments had a positive effect on students' conceptual learning; they demonstrated this positive change by an increase in the number of responses that contained normative ideas close to the scientific explanations. But the findings also indicated that the students did not always show positive progress after they experienced simulations, especially when abstract concepts were introduced. Smetana and Bell (2012) suggest that high-quality support structures (e.g., feedback, training on how to use simulations, and scaffolding) embedded in simulations or provided by instructors, as well as multiple opportunities for students to reflect upon their conceptions, are critical aspects of the successful use of computer simulations. Given the current study's focus on students' individual performance data on assessment tasks, it is suggested that instructional strategies be aligned with assessment tasks, such as providing appropriate training on using simulations, instructors' feedback on student explanations and finding evidence, and more opportunities to reflect on their conceptions in the classroom. The results might be different if the assessment tasks were embedded in the course curriculum or were supported by instructor guidance (Rutten, Joolingen, & van der Veen, 2012). Trundle and Bell (2010) emphasize the connection between computer simulation and pedagogy, especially for promoting conceptual change. Further, they suggest that instructional activities using computer simulations should emphasize a climate of collaboration among students for successful learning. In that regard, the assessment tasks could be used as group tasks to enhance their discourse.

Study limitations and suggested future research are presented below. The study included a relatively small samples of students, a fact that might have compromised the generalization of the results. In an attempt to overcome this limitation, it is necessary to implement the study with a bigger number of students or employ a mixed-methods approach (Creswell, 2014) to provide a more comprehensive examination of the research question.

As students' competency in mathematical modelling plays an important role in learning physics (Redish, 2017), the results might be different if tasks would have been explicitly linked with mathematical modelling of simulated situations with students prompted to demonstrate their competency of mathematical modelling. Also, studies have found that computer simulations are more effective in group or whole-class settings to promote conceptual change (Trundle & Bell, 2010). Future research is required to examine the effectiveness of using the assessment tasks in group or whole-class settings. The study included two tasks to examine students' conceptions in physics; thus, generalizing the results to other topics requires additional research.

REFERENCES

- Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex. *Cell Biology Education*, 4(4), 262-268. <https://doi.org/10.1187/cbe.05-08-0113>
- Barak, J., Gorodetsky, M., & Chipman, D. (1997). Understanding energy in biology and vitalistic conceptions. *International Journal of Science Education*, 19(1), 21-30. <https://doi.org/10.1080/0950069970190102>

- Bell, R. L., & Trundle, K. C. (2008). The use of a computer simulation to promote scientific conceptions of moon phases. *Journal of Research in Science Teaching*, 45(3), 346-372. <https://doi.org/10.1002/tea.20227>
- Black, P., & Solomon, J. (1983). Life world and science world: Pupils' ideas about energy. In G. Marx (Ed.), *Entropy in the school*. Proceedings of the 6th Danube Seminar on Physics Education (pp. 43-455). Budapest: Roland Eoetvoes Physical Society.
- Black, P., & William, D. (1998). *Inside the Black Box: raising standards through classroom assessment*. London: School of Education, King's College.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Bonham, S. W., Deardorff, D. L., & Beichner, R. J. (2003). Comparison of student performance using web and paper-based homework in college-level physics. *Journal of Research in science teaching*, 40(10), 1050-1071. <https://doi.org/10.1002/tea.10120>
- Boyes, E., & Stanisstreet, M. (1990). Misunderstandings of "law" and "conservation": A study of pupils' meanings for these terms. *School Science Review*, 72(258), 51-57.
- Bransford, J. D., Brown, A., & Cocking, R. (Eds.). (1999). *How people learn: Mind brain, experience and school*. Washington, DC: National Academy Press.
- Carey, S. (2009). *The origin of concepts*. Oxford, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367638.001.0001>
- Cheng, K., Thacker, B. A., Cardenas, R. L., & Crouch, C. (2004). Using an online homework system enhances students' learning of physics concepts in an introductory physics course. *American Journal of Physics*, 72(11), 1447-1453. <https://doi.org/10.1119/1.1768555>
- Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: why some misconceptions are robust. *The journal of the learning science*, 14(2), 161-199. https://doi.org/10.1207/s15327809jls1402_1
- Creswell, J. W. (2014). *Research design qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: SAGE Publications.
- Crisp, V., & Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: the PePCAA project. *Computers & Education*, 50(4), 1509-1526. <https://doi.org/10.1016/j.compedu.2007.02.004>
- de Jong, T., & van Joolingen, W. R. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research*, 68(2), 179-201. <https://doi.org/10.3102/00346543068002179>
- de Klerk, S., Veldkamp, B. P., & Eggen, T. J. H. M. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computer & Education*, 85, 23-34. <https://doi.org/10.1016/j.compedu.2014.12.020>
- diSessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10(2-3), 105-225. <https://doi.org/10.1080/07370008.1985.9649008>
- diSessa, A. A. (2002). Why "conceptual ecology" is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice*. Dordrecht, The Netherlands: Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47637-1_2
- diSessa, A.A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, 28, 843-900. https://doi.org/10.1207/s15516709cog2806_1
- Dori, Y. J., & Barak, M. (2001). Virtual and physical molecular modeling: Fostering model perception and spatial understanding. *Educational Technology and Society* 4(1), 61-74.
- Driver, R., & Warrington, L. (1985). Students' use of the principle of energy conservation in problem situations. *Physic Education*, 20, 171-176. <https://doi.org/10.1088/0031-9120/20/4/308>
- Duit, R. (2012). *Towards a learning progression of energy*. Paper presented at the annual meeting of the National Association for Research in Science Teaching (NARST), Indianapolis, IN.
- Eryilmaz, A. (2010). Development and Application of Three-Tier Heat and Temperature Test: Sample of Bachelor and Graduate Students. *Eurasian Journal of Educational Research*, 40, 53-76.
- Geban, O., Askar, P., & Ozkan, I. (1992). Effects of computer simulations and problem solving approaches on high school students. *Journal of Educational Research*, 86(1), 5-10. <https://doi.org/10.1080/00220671.1992.9941821>
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: number and the animate-inanimate distinction as examples. *Cognitive Science*, 14(1), 79-106. https://doi.org/10.1207/s15516709cog1401_5

- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333-2351. <https://doi.org/10.1016/j.compedu.2011.06.004>
- Goldring, H., & Osborne, J. (1994). Students' difficulties with energy and related concepts. *Physics Education*, 29(1), 26 - 31. <https://doi.org/10.1088/0031-9120/29/1/006>
- Gunstone, R. F., & White, R. T. (1981). Understanding of gravity. *Science Education*, 65(3), 291-299. <https://doi.org/10.1002/sce.3730650308>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 989-1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Hammer, D. (1996). More than misconceptions: multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics*, 64(10), 1316-1325. <https://doi.org/10.1119/1.18376>
- Huppert, J., Lomask, S.M., & Lazarowitz, R. (2002). Computer simulations in the high school: Students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education*, 24(8), 803-821. <https://doi.org/10.1080/09500690110049150>
- Hynd, C., Alvermann, D., & Qian, G. (1997). Preservice elementary school teachers' conceptual change about projectile motion: refutation text, demonstration, affective factors, and relevance. *Science Education*, 81(1), 1-27. [https://doi.org/10.1002/\(SICI\)1098-237X\(199701\)81:1<1::AID-SCE1>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1098-237X(199701)81:1<1::AID-SCE1>3.0.CO;2-M)
- Ioannides, C., & Vosniadou, S. (2002). The changing meaning of force. *Cognitive Science Quarterly*, 2(1), 5-61.
- Jimoyiannis, A., & Komis, V. (2001). Computer Simulations in Physics Teaching and Learning: A Case Study on Students' Understanding of Trajectory Motion. *Journal of Computers & Education*, 36, 183-204. [https://doi.org/10.1016/S0360-1315\(00\)00059-2](https://doi.org/10.1016/S0360-1315(00)00059-2)
- Linacre, J. M. (2014). *WINSTEPS (version 3.81)* [Computer program]. Retrieved from <http://www.winsteps.com>
- Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, 87(4), 517-538. <https://doi.org/10.1002/sce.10086>
- Liu, X. (2010). *Essentials of Science Classroom Assessment*. Thousand Oaks, CA: SAGE. <https://doi.org/10.4135/9781483349442>
- Mackey, J. (2009). Virtual learning and real communities: online professional development for teachers. In E. Stacey, & P. Gerbic (Eds.), *Effective blended learning practices: evidence-based perspectives in ICT-facilitated education* (pp. 163-181). Hershey: Information Science Reference. <https://doi.org/10.4018/978-1-60566-296-1.ch009>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149 - 174. <https://doi.org/10.1007/BF02296272>
- McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299-323). Hillsdale, NJ: Laurence Erlbaum Associates.
- Moharreri, K., Ha, M., & Nehm, R. (2014). EvoGrader: an online formative assessment tool for automatically evaluating written evolutionary explanations. *Evolution Education and Outreach*, 7(1), 1-14. <https://doi.org/10.1186/s12052-014-0015-2>
- Navarro, P., & Shoemaker, J. (2000). Performance and perceptions of distance learners in cyberspace. *American Journal of Distance Education*, 14(2), 15-35. <https://doi.org/10.1080/08923640009527052>
- Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, 48(3), 237-256. <https://doi.org/10.1002/tea.20400>
- Nguyen, T. (2015). The effectiveness of online learning: Beyond no significant difference and future horizons. *MERLOT Journal of Online Learning and Teaching*, 11(2), 309-319.
- Ozdemir, G. & Clark, D. (2009). Knowledge structure coherence in Turkish students' understanding of force. *Journal of Research in Science Teaching*, 46(5), 570-596. <https://doi.org/10.1002/tea.20290>
- Pachler, N., Daly, C., Mor, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education*, 54, 715-721. <https://doi.org/10.1016/j.compedu.2009.09.032>
- Park, M., & Liu, X. (2016). Assessing understanding of the energy concept in difference science disciplines. *Science Education*, 100(3), 483-516. <https://doi.org/10.1002/sce.21211>
- Park, M., & Liu, X. (2019). An investigation of item difficulties in energy aspects across biology, chemistry, environmental science, and physics. *Research in Science Education*. Published online first. <https://doi.org/10.1007/s11165-019-9819-y>

- Park, M., Liu, X., Smith, E., & Waight, N. (2017). The effect of computer models as formative assessment on student understanding of the nature of models. *Chemistry Education Research and Practice*, 18, 572-581. <http://doi.org/10.1039/c7rp00018a>
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco, CA: Jossey-Bass Publishers.
- Penn, J., Nedeff, V. M., & Gozdzik, G. (2000). Organic chemistry and the internet: a web-based approach to homework and testing using the WE_LEARN System. *Journal of Chemical Education*, 77(2), 227-231. <https://doi.org/10.1021/ed077p227>
- Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., & Wieman, C. (2006). PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher*, 44, 18-23. <https://doi.org/10.1119/1.2150754>
- Peterson, R. F., Treagust, D.F., & Garnet, P. (1989). Development and application of diagnostic instrument to evaluate grade-11 and -12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching*, 26(4), 301-314. <https://doi.org/10.1002/tea.3660260404>
- Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching*, 49(3), 363-393. <https://doi.org/10.1002/tea.21005>
- Redish, E. F. (2017). Analysing the competency of mathematical modelling in physics. In: Greczyło, T., & Dębowska, E., (Eds.), *Key Competences in Physics Teaching and Learning* (pp. 25-40). Chum: Springer International Publishing. https://doi.org/10.1007/978-3-319-44887-9_3
- Richards-Babb, M., Drelick, J., Henry, Z., & Robertson-Honecker, J. (2011). Online homework, help or hindrance? What students think and how they perform. *Journal of College Science Teaching*, 40(4), 81-93.
- Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136-153. <https://doi.org/10.1016/j.compedu.2011.07.017>
- Smetana, L., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education*, 34(9), 1337-1370. <https://doi.org/10.1080/09500693.2011.605182>
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: A meta-analysis. *Review of Educational Research*, 69, 21-51. <https://doi.org/10.3102/00346543069001021>
- Srisawasdi, N., & Kroothkeaw, S. (2014). Supporting students' conceptual learning and retention of light refraction concepts by simulation-based inquiry with dual-situated learning model. *Journal of Computers in Education*, 1(1), 49-79. <https://doi.org/10.1007/s40692-014-0005-y>
- Srisawasdi, N., & Panjaburee, P. (2015) Exploring effectiveness of simulation-based inquiry learning in science with integration of formative assessment, *Journal of Computers in Education*, 2(3), 323-352. <https://doi.org/10.1007/s40692-015-0037-y>
- Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: a review of the research. *Review of Educational Research*, 76(1), 93-135. <https://doi.org/10.3102/00346543076001093>
- Tao, P-K., & Gunstone, R. (1999). The process of conceptual change in force and motion during computer-supported physics instruction. *Journal of Research in Science Teaching*, 36(7), 859-882. [https://doi.org/10.1002/\(SICI\)1098-2736\(199909\)36:7<859::AID-TEA7>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1098-2736(199909)36:7<859::AID-TEA7>3.0.CO;2-J)
- Tatar, E., & Oktay, M. (2007). Students' misunderstandings about the energy conservation principle: a general view to studies in literature. *International Journal of Environmental & Science Education*, 2(3), 79-81.
- Treagust, D. F. (1998). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159-169. <https://doi.org/10.1080/0950069880100204>
- Treagust, D. F., Chittleborough, G., & Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education*, 24(4), 357-368. <https://doi.org/10.1080/09500690110066485>
- Trundle, K. C., & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: A quasi-experimental study. *Computers & Education*, 54(4), 1078-1088. <https://doi.org/10.1016/j.compedu.2009.10.012>

- Winn, W., Stahr, F., Sarason, C., Fruland, R., Oppenheimer, P., & Lee, Y. (2006). Learning oceanography from a computer simulation compared with direct experience at sea. *Journal of Research in Science Teaching*, 43(1), 25-42. <https://doi.org/10.1002/tea.20097>
- Zacharia, Z., & 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(6), 618-629. <https://doi.org/10.1119/1.1566427>

<http://www.ejmste.com>