Journal of Research in STEM Education

Vol 2, No 1, July 2016, PP 39-58

RESEARCH REPORT

Exploring Student Understanding of Force and Motion Using a Simulation-Based Performance Assessment

Jessica Gale [a1], Jayma Koval[a], Stefanie Wind[b], Mike Ryan[a], Marion Usselman[a]

[a]*Georgia Institute of Technology, USA;* [b]*University of Alabama, USA*

Abstract: *Performance assessment (PA) has been increasingly advocated as a method for measuring students' conceptual understanding of scientific phenomena. In this study, we describe preliminary findings of a simulation-based PA utilized to measure 8th grade students' understanding of physical science concepts taught via an experimental problem-based curriculum, SLIDER (Science Learning Integrating Design Engineering and Robotics). In SLIDER, students use LEGO robotics to complete a series of investigations and engineering design challenges designed to deepen their understanding of key force and motion concepts (net force, acceleration, friction, balanced forces, and inertia). The simulation-based performance assessment consisted of 4 tasks in which students engaged with video simulations illustrating physical science concepts aligned to the SLIDER curriculum. The performance assessment was administered to a stratified sample of 8th grade students (N=24) in one school prior to and following implementation of the SLIDER curriculum. In addition to providing an illustration of the use of simulation-based performance assessment in the context of design-based implementation research (DBIR), the results of the study indicate preliminary evidence of student learning over the course of curriculum implementation.*

Keywords: Physical Science, Performance Assessment, Middle Grades Science

The need to produce more STEM graduates to maintain the national security and economic future of United The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) offer the following definition of performance assessments (PA): "assessments for which the test taker actually demonstrates the skills the test is intended to measure by doing tasks that require those skills" (p. 221). PA has been promoted as providing more direct or authentic measurement of student achievement than selected-response formats, such as multiple-choice assessments (Lane & Stone, 2006). PAs have been touted as essential indicators of student mastery of science content and skills that can serve as both formative and summative assessments (Lane & Stone, 2006). Lane and Stone argue, "to fully capture the essence of scientific inquiry requires the use of hands-on performance tasks that may be extended over a number of days" (p. 388). This perspective is echoed by the Committee on Developing Assessments of Science Proficiency in K-12 for the Next-Generation Science Standards (NGSS) recommendation that assessment tasks "should include—as a significant and visible aspect of the assessment—multiple, performance-based questions" (National Research Council, 2014, p. 7).

PA has been described as a useful method for assessing conceptual development and documenting students' alternative conceptions (i.e., misconceptions, naïve/intuitive theories). PA methods used in science education include tasks asking students to interact with physical stimuli and explain scientific phenomena (e.g. McCloskey, 1983) or draw pictures depicting their conceptual understanding (Vosniadou & Brewer, 1994). Despite considerable attention to PA, implementation is often limited by practical constraints related to time,

[1] Corresponding author. *Georgia Institute of Technology, Center for Education Integrating Science, Mathematics, and Computing (CEISMC), USA*, E-mail: jessica.gale@ceismc.gatech.edu

Gale, J., Koval, J., Wind, S., Ryan, M., & Usselman, M. (2016). Exploring student understanding of force and motion using a simulation-based performance assessment. *Journal of Research in STEM Education*, *2*(1), 39-58

ISSN:2149-8504 (online)

resources, and costs. Given these limitations, there are few examples of research utilizing performance assessments to measure science students' conceptual understanding over the course of curricular interventions.

Simulation-based Performance Assessment

The *Standards* note that simulation-based assessment formats may be especially appropriate in contexts where "actual task performance might be costly or dangerous" (AERA, APA, & NCME, 2014, p. 78). Similarly, the National Research Council (NRC) report Knowing What Students Know asserts "technology is making it possible to assess a much wider range of important cognitive competencies than was previously possible. Computer-enhanced assessments can aid in the assessment of problem-solving skills by presenting complex, realistic, open-ended problems…" (Pellegrino, Chudowsky, & Glaser, 2001, p. 266). Thus, simulation-based assessments offer a potential compromise, allowing for representation of scientific phenomena without the constraints and limitations inherent in performance assessments that involve student interaction with physical demonstrations or stimuli.

As efforts to enhance science education have employed innovative computer-based activities and simulations, researchers have begun to explore creative approaches to utilizing simulations for assessment (Thompson Tutwiler, Metcalf, Kamarainen, Grotzer, & Dede, 2016; White & Frederiksen, 2000). A number of projects have experimented with computer-based tasks intended to document and track learners' developing understandings or knowledge representations, such as through the creation of concept maps (O'Neil and Klein, 1997) or their development of persuasive arguments (Mislevy, Steinberg, Almond, Haertel, and Penuel, 2000). Similarly, the EcoXPT project (Thompson et al., 2016) has adopted a blended assessment strategy, with traditional assessments complemented by the analyses of log file data generated from student engagement within a multi-user virtual environment.

This study illustrates the use of a set of iteratively developed simulation-based performance assessment (PA) tasks within the context of a design-based implementation research (DBIR) project. Specifically, we describe data collected from the administration of four simulation-based PA tasks designed to assess 8th grade students' understanding of force and motion concepts following implementation of an experimental problem-based curriculum. Through illustrative examples and the analysis of student responses to PA tasks administered prior to and following the curriculum implementation, the study provides illustrative results from a sample of (N=24) of 8th grade students.

Methodology

This section describes the curricular context in which the assessment was conducted, the sample of students that participated in this study, and the simulation-based PA tasks.

Curricular Context: The SLIDER Project

SLIDER is an NSF-funded DRK-12 project examining the use of design and engineering, through LEGO robotics, in the context of 8th grade physical science classrooms. The SLIDER curriculum, which is comprised of two 5-week units, was iteratively developed over a three-year period within diverse school contexts, ranging from affluent, high-achieving suburban classrooms to relatively low-proficiency, low-income rural schools (Usselman & Ryan, 2014). SLIDER features contextualized design challenges intended to facilitated student learning of key physical science concepts. In SLIDER Unit 1, students apply their understanding of energy concepts (e.g. energy transfer, potential and kinetic energy) to engineer a solution to a traffic problem scenario - increased accidents at a dangerous intersection in a fictional town. SLIDER Unit 2 focuses on force and motion concepts (net force, balanced forces, acceleration, inertia) and culminates in a design challenge in which students use LEGO Mindstorms™ kits to design and test an automatic braking system for a robotic truck.

For additional information about the SLIDER project and access to SLIDER curriculum materials visit https://slider.gatech.edu/.

Participants

The PA was administered to 24 eighth grade physical science students taught by a teacher implementing the SLIDER curriculum at a suburban middle school in the southeastern United States during the 2014-15 school year. Students were sampled from this particular teacher's classes because the teacher exhibited high fidelity of implementation of the curriculum relative to other SLIDER teachers. A mixed-methods sampling strategy was utilized in order to include students representing a range of achievement levels (Teddlie & Yu, 2007). Sampling began with analysis of student performance on multiple-choice items in the SLIDER Unit 2 pre-assessment. Using the dichotomous Rasch model (see Engelhard, 2013) to estimate student achievement, students were classified into achievement-leveled groups based on performance on the SLIDER Unit 2 pre-assessment (high, medium, and low). The second stage of the sampling procedure utilized reputational case selection (Goetz & LeCompte, 1984). The teacher was presented with a matrix of student names grouped by class period and achievement level and asked to recommend 24 students (eight students from each achievement level column) who had consistent attendance and had actively participated in SLIDER activities. The teacher was not informed that the three columns in the matrix represented student grouping based on achievement.

The SLIDER Simulation-Based Performance Assessment Tasks

The project utilized a multilevel approach to assessment (Ruiz-Primo, Shavelson, Hamilton, Klein, 2002; Hickey & Zuiker, 2012) in order to investigate student understanding of force and motion concepts within the SLIDER curriculum. In this approach, a variety of assessments are used based on their proximity to the curriculum being implemented. Student work or artifacts generated through students' interaction with the curriculum are considered immediate assessments. Close assessments align with the specific content and activities within the curriculum. Proximal assessments measure the acquisition of knowledge and skills relevant to the curriculum, but the topics or context of the assessment tasks can be different. Distal assessments, such as standardized tests, typically represent state or national standards in a specific discipline. Accordingly, the PA tasks described below serve as a proximal assessment that complements a set of other immediate- and close-level assessments imbedded within the curriculum and additional relatively distal assessments including standardized multiple-choice items. As proximal-level assessments, the tasks presented problem-solving scenarios that aligned to the same physical science concepts as the curriculum but differed in terms of context and, in some cases, difficulty. For example, within the SLIDER curriculum, students are asked to reason about force and motion in the context of automobile collisions (e.g. trucks hitting cars). In the PA, students are asked to transfer the knowledge they learned through SLIDER to answer different types of questions in a different context (figures pushing or pulling boxes).

The PA instrument includes four tasks, developed in collaboration with the SLIDER curriculum team to assess student understanding of major concepts addressed within the curriculum: net force, acceleration, friction, balanced forces, and inertia. The tasks were developed by adapting simulations from the University of Boulder PhET Interactive simulations (available online at: https://phet.colorado.edu/). Video-editing software was used to create short video clips portraying the selected PhET simulations for each task. Each of the four PA tasks is described below. (See Gale, Wind, Koval, Dagosta, Ryan, and Usselman, 2016 for additional details about the development and administration of the PA tasks).

Task 1: Net Force. Task 1, depicted in Figure 1, asked students to describe the net force represented in three tug-of-war scenarios. The researcher introduced the task by explaining that the tug-of war in the task was between two teams, and that figures from each team would pull the rope to move the cart over to their side. Students were told to disregard friction, gravity and the force from the ground (e.g. normal force) and that they should only consider forces from the figures pulling the rope. The task proceeded with three scenarios in which

ISSN:2149-8504 (online)

students were shown illustrations and asked to indicate whether there was a net force (e.g. "If we have four people of equal strength on each side, will there be a net force when the tug-of-war begins?"). When students predicted that there would be a net force, they were shown two arrows, a large arrow and a small arrow, and asked to choose and place it the illustration to show the net force. Students then watched a video simulation of the scenario and compared the result to their prediction.

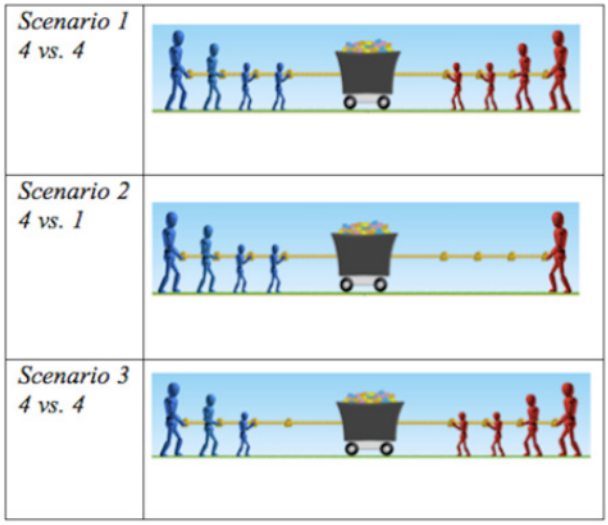


Figure 1. Task 1: Net Force

Task 2: An Object in Motion Task 2, depicted in Figure 2, assessed students' understanding of net force using a simulation in which a figure pushes a box along a surface that they are told has a medium amount of friction. The speed of the figure increases as it pushes the box until the point is reached where the figure can no longer keep up with the box and falls away. The box continues to move forward but the speed decreases and eventually the box comes to a complete stop. After viewing the full simulation video, the researcher plays the video a second time, pausing to ask students to identify and explain the direction of the net force at three time-points: when the figure pushed the box as the speed was increasing; after the figure fell away from the box and the speed was decreasing; and once the box came to a complete stop. At each time-point students were asked, "Is there a net force?" If they answered yes, they were asked to select either a large or a small arrow and place it on an illustration of the tug-of-war event to show the direction of the net force and to explain their placement of the arrow ("Tell me why you placed the arrow the way you did to describe the net force").

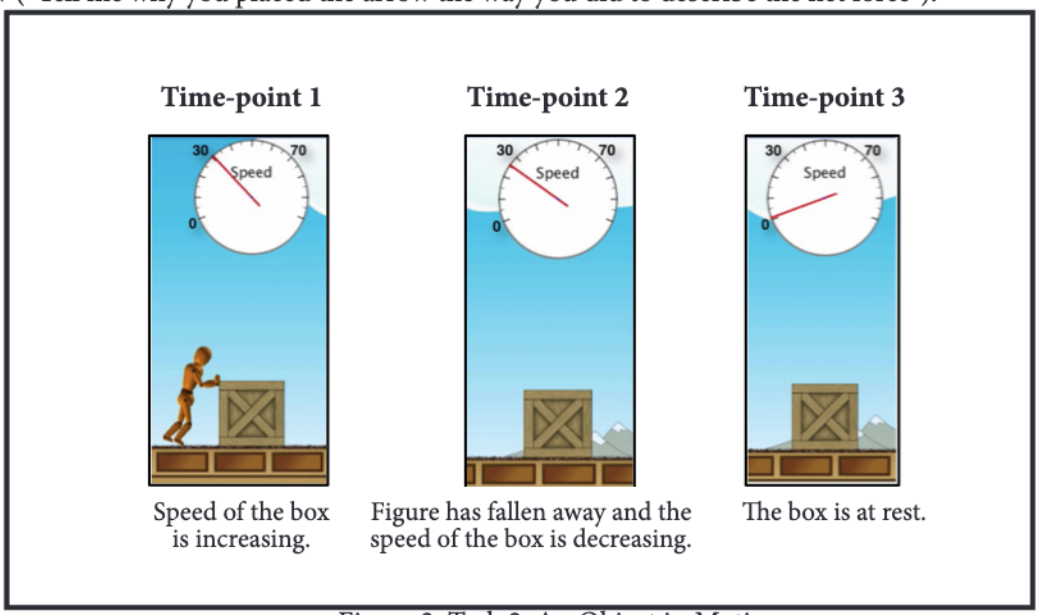


Figure 2. Task 2: An Object in Motion

Task 3: Balanced Forces. In Task 3, depicted in Figure 3, students considered a scenario in which they were asked to explain how a constant speed could be achieved. In the video simulation, they watched a figure push a box until it reached a speed of 70. Students learned that the figure was pushing with 250 N of applied force and the force of friction was 125 N. When the box reached the speed of 70, the researcher paused the video, presented a picture of the same moment and asked, "Let's say the figure wants to keep the speed at 70. What could the figure do to make that happen?" Additional probing questions were used, as necessary, to elicit student explanations. Specifically, researchers sought to determine whether students held the common misconception that balancing forces would cause the object to stop. Therefore, if students responded that the figure should push with more than 125N of force, the researcher probed with the question, "what do you think would happen if the figure pushed with 125N?"

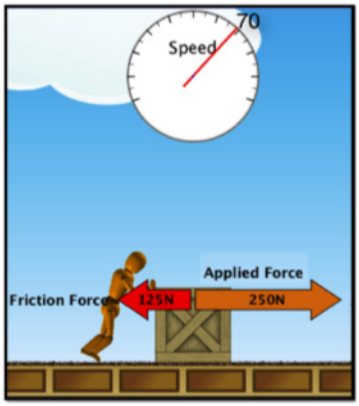


Figure 3. Task 3: Balanced Forces

Task 4: Inertia. Task 4, depicted in Figure 4, was designed to reveal students' understanding of inertia. First, students watched the figure push a box using 300N of force and use a stopwatch to measure how many seconds it took for the figure to push the box from a resting position to reach a speed of 70. In the second half of the simulation a second box was stacked on top of the first and the figure again used 300N of force to push the box from rest to a speed of 70. Before watching the simulation students were asked predict how long they thought it would take and why ("How many seconds do you think it will take for the boxes to reach a speed of 70…Why do you predict____ seconds?"). Students then used a stopwatch to measure how long it took for the figure to push two boxes to the target speed of 70. Students were then asked to explain why it took so much longer for the figure to push two boxes ("With one box, it took ____ seconds. With two boxes, it took ____ seconds. Why do you think that happened?") If students didn't mention inertia independently in their answer, they were prompted to describe the event in terms of inertia ("What can you tell me about inertia that might explain why this happened?").

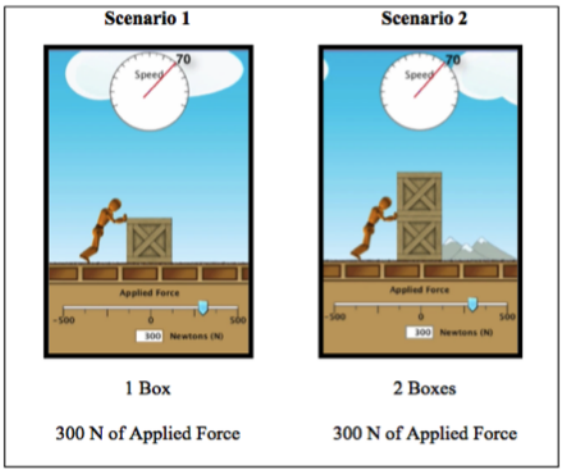


Figure 4. Task 4: Inertia

Performance Task Administration

Task administration followed a protocol with a format similar to a semi-structured interview. The PA was conducted by the same member of the research team just prior to the implementation of SLIDER Unit 1 (Pre-PA) and approximately 3 months later (Post-PA), immediately following implementation of the SLIDER curriculum's second unit. This researcher had visited the participating classroom several times prior to the PA task administration, so students were accustomed to her presence and generally comfortable speaking with her. All performance assessment sessions were videotaped. A second researcher was present during PA administration to operate video recording equipment and take notes on student responses for each task. The PA took approximately 15 minutes per student for each administration and was conducted in a quiet area near the science classroom.

Data Analysis

Pre- and post- responses for each task were analyzed for each of the twenty-four participating students. Because student responses for PA Task 1 were limited to answering "yes" or "no" to the prompt "Is there a net force?", and to placing an arrow to indicate net force, Task 1 data was compiled from data sheets completed by researchers during task administration. Video recordings for tasks 2-4 were transcribed for analysis. Using the NVIVO software program, all student responses were coded by two members of the research team, including the researcher who administered the performance assessment. All student responses (both pre- and post-) were compiled in an NVIVO project file such that coders were blind to whether a student response was from the pre- or post-PA administration. Coding followed a protocol coding process (Saldana, 2013) wherein student responses were evaluated using a task-specific rubric iteratively developed by the research team. The rubric included two types of codes: holistic codes and explanation codes. *Holistic* codes, defined at four levels of understanding for each task, were utilized to describe the degree to which student responses were indicative of accurate conceptual understanding of targeted science concepts. Although rubrics were task specific, they generally defined a similar progression of conceptual understanding: "incorrect" responses indicative of alternative understandings inconsistent with accepted scientific understandings of force and motion concepts were coded at Level 1; "correct" responses consistent with accepted scientific understandings were coded at Level 2; and responses that were both "correct" and included an explanation that accurately referred to or applied a relevant force or motion concept were coded at Level 3. Following coding, differences between pre- and post rubric scores for each task were investigated using Wilcoxon signed-rank tests (Corder & Forman, 2014). Further analysis of student responses included the application of *Explanation* codes, which categorized the explanations and predictions students provided within the tasks and indicated whether students arrived at their ultimate responses independently or through follow-up questions from the researcher, which we refer to as "prompting". Task rubrics (see Appendix) were revised with input from the SLIDER research team following a first round of coding. Following a second round of coding, coder comparison queries indicated 94% agreement between coders across tasks. Remaining coding discrepancies were resolved through discussion between coders.

Results

This section presents results and illustrative examples for SLIDER's simulation-based performance assessment tasks, beginning with descriptive results for the introductory Task 1 and followed by results and illustrative examples of student responses for Tasks 2-4.

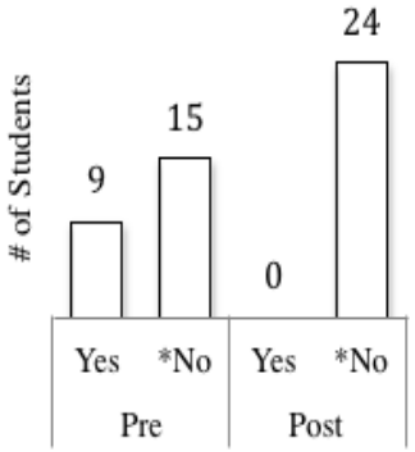
Task 1

Student responses to the Task 1 prompt, "Is there a net force?" and their ability to correctly place an arrow indicating the direction and magnitude of the net force, suggest subtle differences between pre-and post response patterns. As indicated in Figure 5, on the pre-PA, nine of the 24 students incorrectly stated that there was a net force in Scenario One. Asked to describe the net force, five of these students were unable to give a response or said "I don't know" and four students stated that the net force is "the same on each side", suggesting potential confusion between the vocabulary "net force" and "force". For Scenario Two, nearly all students responded correctly to both prompts at both pre- and post-PA. For Scenario Three, at both pre- and post-PA all students correctly affirmed the net force and correctly indicated the direction of the net force; however, there was an increase in the number of students who selected the small arrow to correctly indicate the magnitude of the net force from pre- to post-PA.

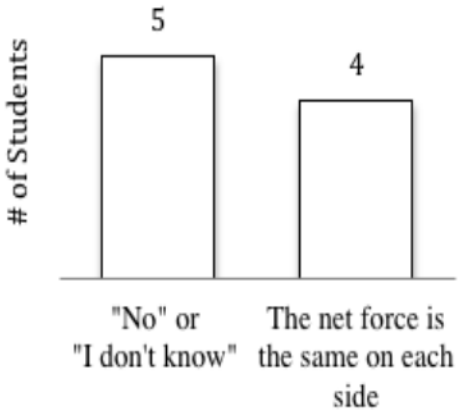
That even students who responded incorrectly on scenario one were able to correctly state whether there was a net force in scenarios two and three suggests that students who began the task with a lack of understanding of net force may have learned the basic concept over the course of the task. Given the simplicity of the task and that students were shown simulation videos illustrating the outcomes for each tug-of-war scenario after giving their response, it is also possible that students simply inferred the basic meaning of "net force" rather than developing an accurate understanding of the concept. Thus, "correct" answers to the yes/no questions in scenarios two and three do not necessarily indicate fully developed conceptual understanding.

ISSN:2149-8504 (online)

Scenario One

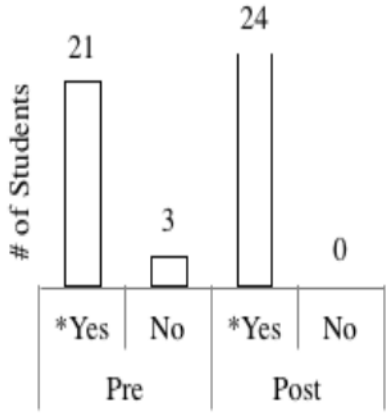


"Will there be a net force?"

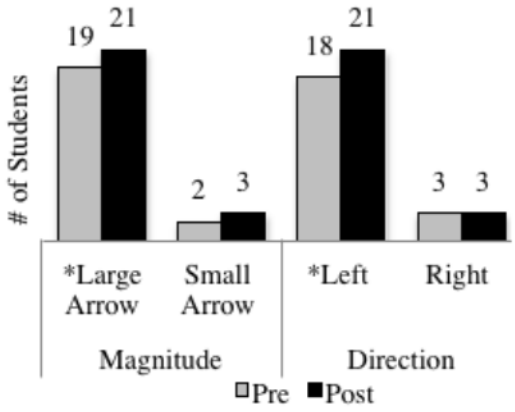


(If Yes)" Can you describe the net force."

Scenario Two

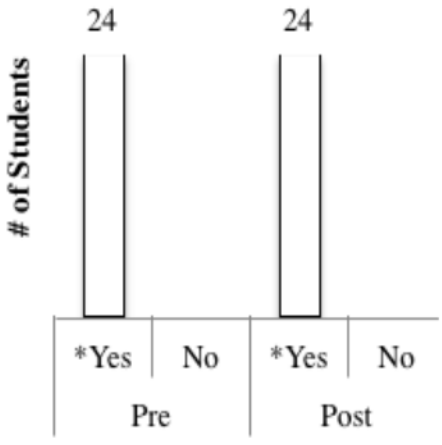


"Will there be a net force?"

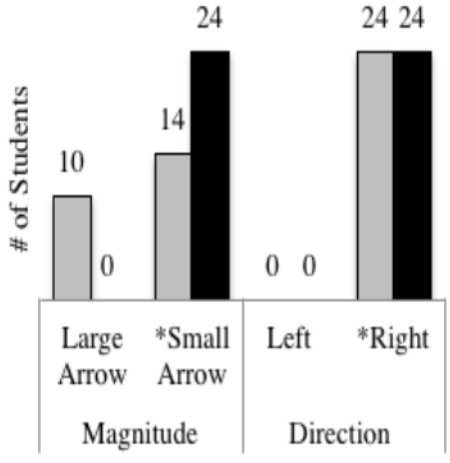


(If Yes) "Choose an arrow and place it to show the net force."

Scenario Three



"Will there be a net force?"



(If Yes) "Choose an arrow and place it to show the net force."

In addition to assessing students' understanding of net force, Task 1 was intended to serve as an introduction to the simulation-based performance task format and provide a mastery experience for students presenting more conceptually difficult tasks that would require students to provide explanations of force and motion phenomena depicted in simulations. The ease with which students responded to the prompts suggests that Task 1 was successful in this regard.

Task 2

Recall that in Task 2, students viewed a simulation that depicted a box in various states of motion at three time points. Students were asked at each time point whether there was a net force acting on the box, to indicate the direction of the net force using an arrow, and to explain why they placed the arrow where they did to show the net force.

Figure 6 depicts student-level rubric scores at pre- and post-PA administrations. Prior to SLIDER implementation, 20 of the 24 students gave a Level 1 response, inaccurately stating whether there was a net force and/or indicating the incorrect direction of the net force. Relatively few students provided explanations that referred to applied force and/or friction (Level 2) or compared applied and frictional forces (Level 3). Although three of these students maintained this inaccurate response at post-PA, seventeen students provided scientifically accurate responses following SLIDER and the majority of these students (n=10) progressed from a Level 1 to a Level 3 response in which they not only correctly indicated the net force but also explained their response by explicitly discussing balanced forces or comparing the relevant applied and frictional forces within the simulation scenario. These patterns are consistent with Wilcoxon signed-rank tests showing statistically significant changes in rubric ratings between pre- and post-PA administrations for Task 2 ($Z = -3.93$, $p < .001$).

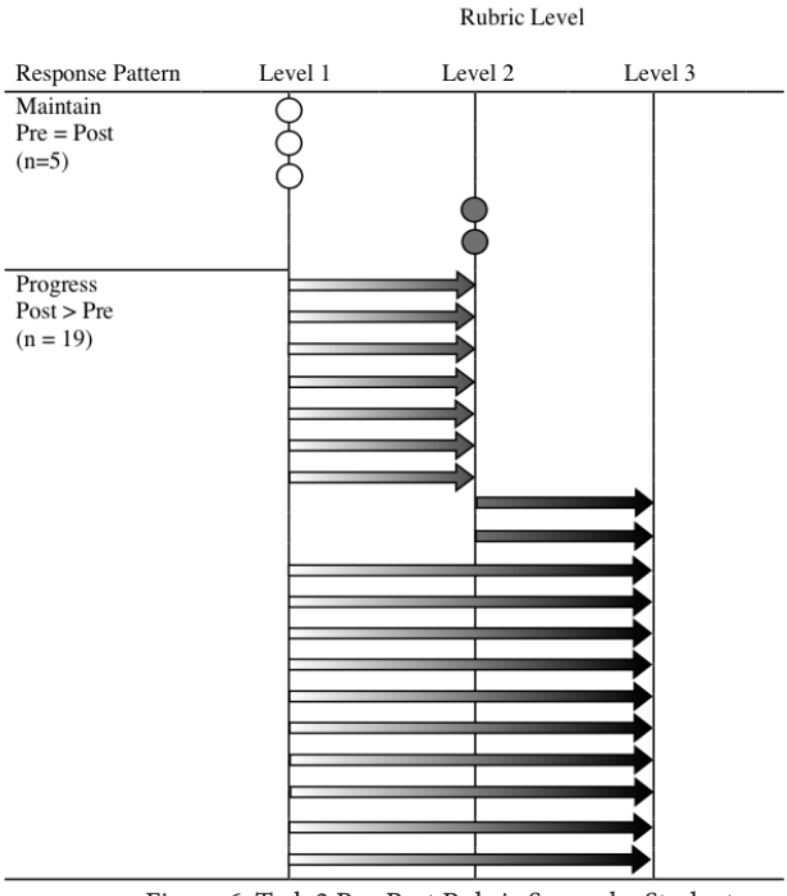


Figure 6. Task 2 Pre-Post Rubric Scores by Student.

Note: Dots represent unchanged rubric scores and arrows represent pre-post changes in rubric level scores for each student. See Appendix A for rubric level definitions.

Figure 7 illustrates the pattern of student responses when asked to explain their responses when the box was moving (Time-points 1 and 2) and when the box was at rest (Time-point 3). Note that because time-points 1 and 2 represent conceptually similar events (the box in motion), student responses at these two time-points were combined for analysis.

Taken together, student responses coded using the holistic and explanation rubrics illustrate a shift in student understanding of the targeted physical science concepts assessed by Task 2. This shift in understanding is further illustrated in the example presented in Table 1, in which the student provides a Level 1 response prior to SLIDER and a Level 3 response following curriculum implementation.

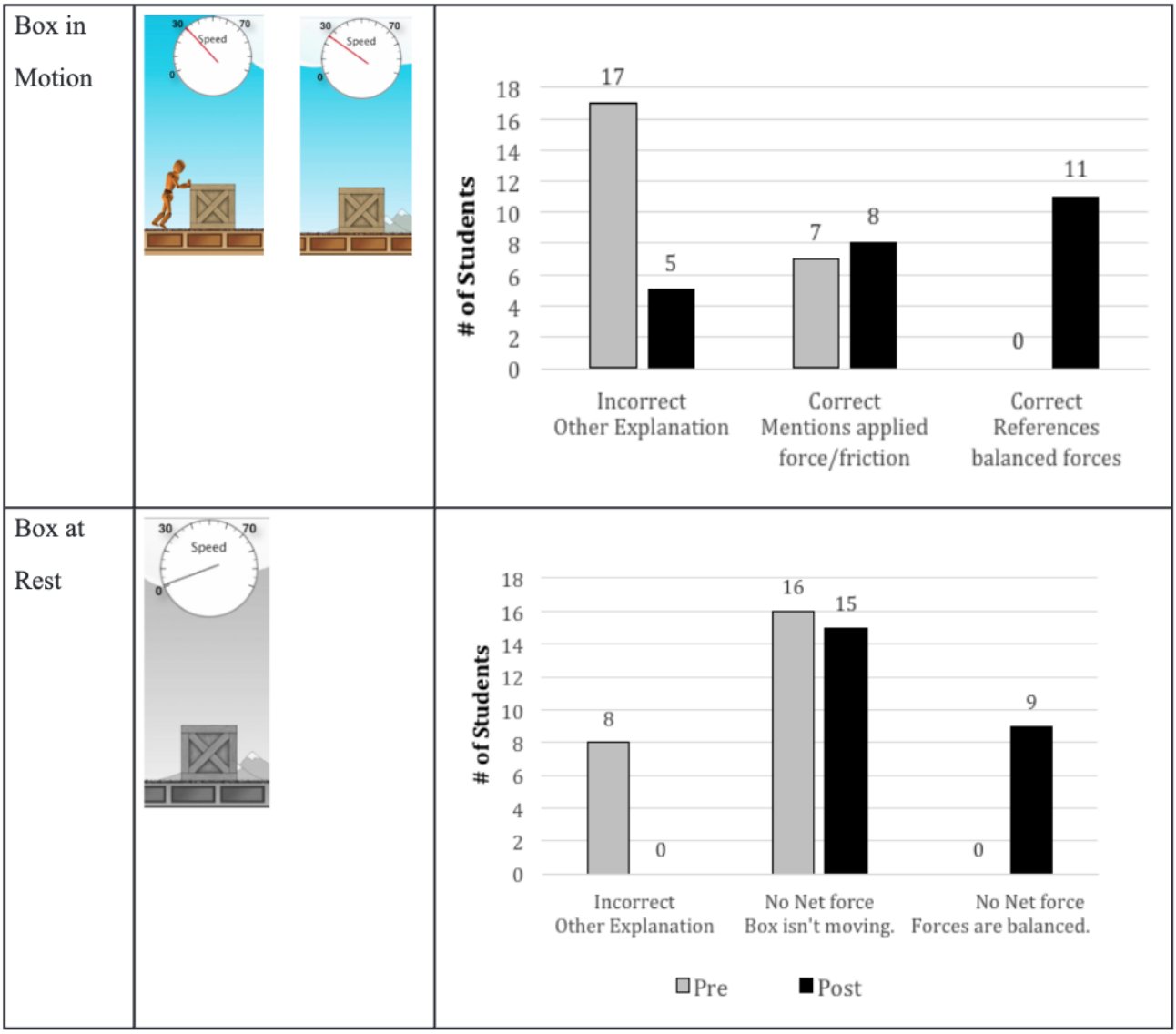


Figure 7. Task 2 Student Explanations

Table 1.
Task 2 Illustrative Example

Pre	Post
Time-point 1: R: Is there a net force acting on the box? S: Yes. R: Please place the arrow on the picture to show the net force. S: (Student places the arrow pointing to the right.) R: Tell me why you placed the arrow there? S: Because the man is pushing the box forward.	Time-point 1: R: Is there a net force acting on the box? S: Yes. R: Please place the arrow on the picture to show the net force. S: (Student places arrow pointing to the right.) R: Tell me why you placed the arrow there? S: Because the man is pushing the box and the amount of force he's using is greater than the amount of friction.
Time-point 2: R: …Is there a net force acting on the box? S: No. R: Tell me why. S: Because there's nothing moving the box in that direction.	Time-point 2: R: Is there a net force acting on the box? S: Yes. R: Please place the arrow on the picture to show the net force. S: (Student points arrow pointing to the left.) R: Tell me why you placed the arrow there? S: Because the man is no longer pushing it and the friction is greater than the force that is pushing it now.
Time-point 3: R: Is there a net force acting on the box? S: No. R: Tell me why. S: Because nothing is pushing the box in the right direction or the left direction.	Time-point 3: R: Is there a net force acting on the box? S: No. R: Tell me why. S: Because the box has stopped moving, there was no more friction affecting it and the box can't move forward because there is no one to push it forward.
[Response scored at Rubric Level 1]	[Response scored at Rubric Level 3]

Note: S = Student, R=Researcher.

Task 3

Recall that Task 3 asked students to reason about how a box being pushed with 250N of applied force could maintain a constant speed. Students answered the question "Let's say the figure wants to keep the speed at 70. What could the figure do to make that happen?" (See Figure 2). Figure 8 illustrates the distribution of students' scores on the holistic rubric for Task 3. These results suggest some development in students' understanding of how balanced forces operate when an object is in motion, with an increase in the number of students who explicitly referred to balanced forces when concluding that the figure should push the box with 125N of force to maintain its speed. At the same time, the persistence of incorrect Level 1 responses and the fact that four students exhibited a regressive response pattern, scoring lower on the holistic rubric at post-test than at pre-test, suggests that this was a particularly difficult task for many students. These patterns are consistent with Wilcoxon signed-rank tests showing a non-significant change in students' holistic rubric scores for Task 3.

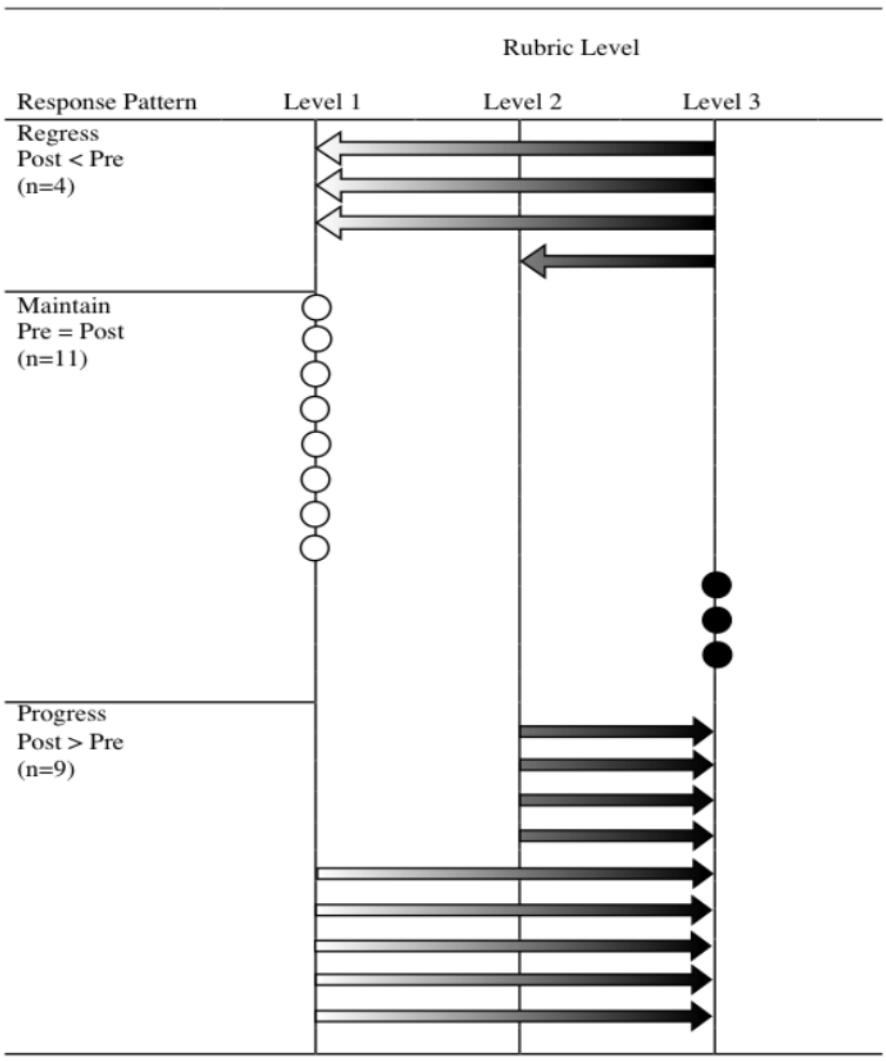


Figure 8. Task 3 Pre-Post Rubric Scores by Student.

Note: Dots represent unchanged rubric scores and arrows represent pre-post changes in rubric level scores for each student. See Appendix A for rubric level definitions.

Figure 9 presents the distribution of student responses to the Task 3 question "What could the figure do to keep the speed at 70?". At both administrations, students who provided an incorrect response were most likely to state that the figure should push with a force that is less than 250N but more than the frictional force of 125N. Further questioning revealed that a number of students providing this response (two at pre-PA and six at post-PA) held the misconception that if the forces were balanced such that the figure pushed with an applied force equal to the frictional force, the box would stop moving, a misconception that is well documented in the science education literature (AAAS, 2010). Figure 9 also illustrates the number of students who arrived at correct responses independently or through prompting at both the pre- and post- administrations of the PA. When students provided incorrect (Level 1) responses, researchers engaged students in further discussion in order to clarify or more fully reveal students' understanding. While the intention of these follow-up questions was not necessarily to lead students to change their answers but rather to clarify students' responses, we did find that, in some cases, students' responses in Task 3 evolved over the course of these discussions. A number of students at both administrations initially provided incorrect responses but arrived at the correct response through discussion; however, students were somewhat more likely to independently provide correct responses following the SLIDER curriculum.

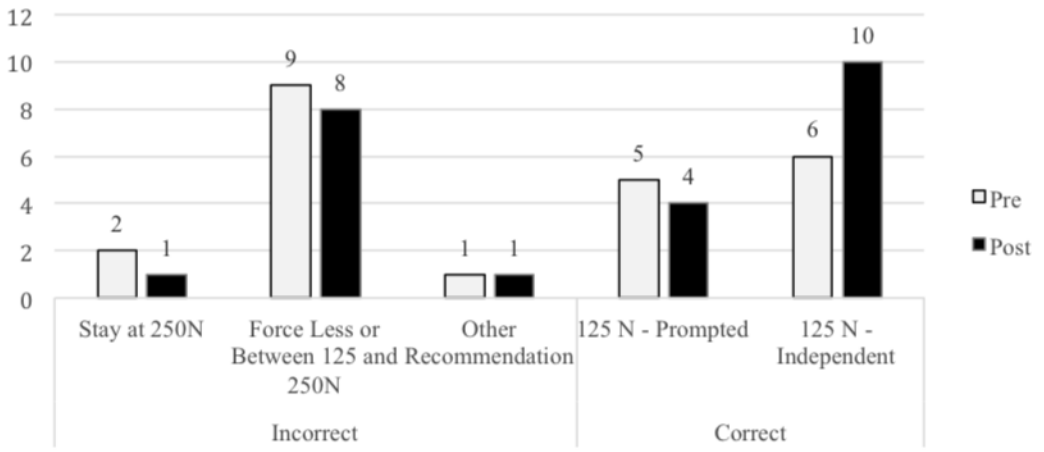


Figure 9. Student Responses to Task 3 Question: What could the figure do to keep speed at 70?

Table 2 presents an illustrative example of one students' Task 3. Prior to engaging with the SLIDER curriculum, the student initially gave a response approximating the scientifically accurate understanding that balancing the force with which the box is pushed and the force of friction would result in a constant speed. However, the student then changes his response, articulating the alternative understanding that balanced forces would cause the box to stop moving. Following SLIDER, the student seems to have revised his understanding to confirm his initial conception that balanced forces would produce a constant speed.

Table 2.

Task 3 Illustrative Example

Pre	Post
R: Let's say the figure wants to keep the speed at 70. What could the figure do to make that happen? S: They would lessen their force a little bit so that the forces would be equal. And then there wouldn't be a net force. But it would keep its speed….No. No. It would just make it go down. He would make his force go down a little, but not all the way to 125, because that would mean the box wouldn't be moving. So maybe to just about 200, or somewhere around there." [Scored at Rubric Level 1]	R: Let's say the figure wants to keep the speed at 70. What could the figure do to make that happen? S: It would cut its force in half because then that would balance out the forces and then it would just keep moving at a constant speed. [Scored at Rubric Level 3]

Note: S = Student, R = Researcher.

Task 4

Recall that Task 4 focused on the concept of inertia and asked students to predict and explain an increase in the time required for the figure to reach a certain speed when pushing two boxes versus one box. Figure 10 illustrates the distribution of students' scores on the holistic coding rubric for Task 4. These holistic coding results suggest a progression in students' understanding of inertia. All but one student provided responses indicating an understanding of inertia on the post-PA and there was an apparent shift in the extent to which students explicitly applied the concept of inertia to explain what they observed in the simulation. These patterns are consistent with Wilcoxon signed-rank tests showing statistically significant changes in rubric ratings between pre- and post-PA administrations for Task 4 ($Z = -3.72$, $p < .001$).

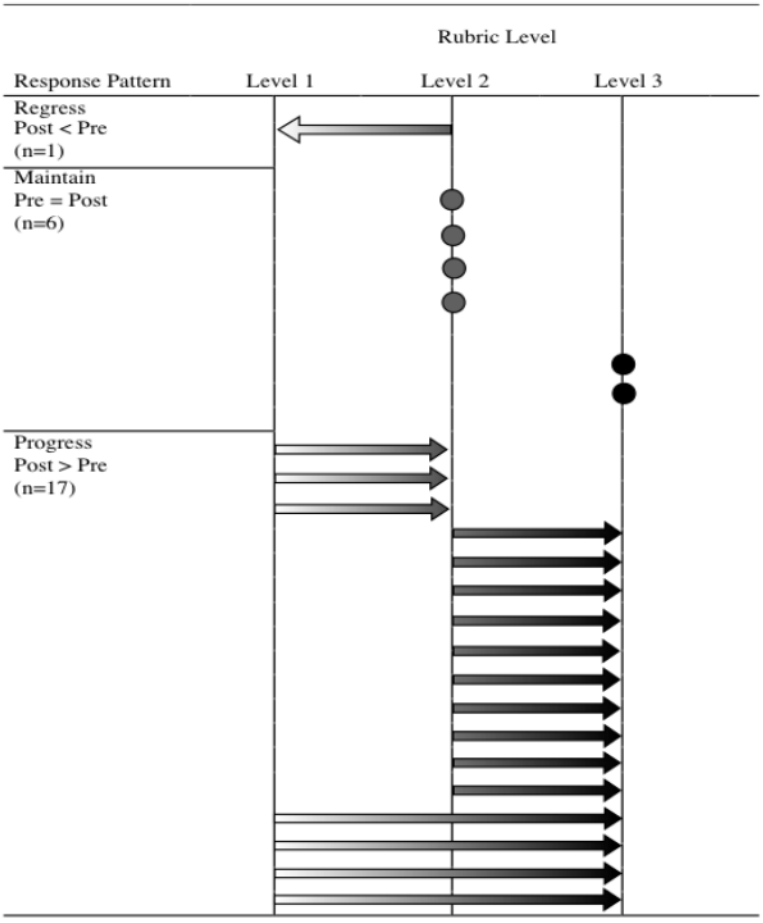
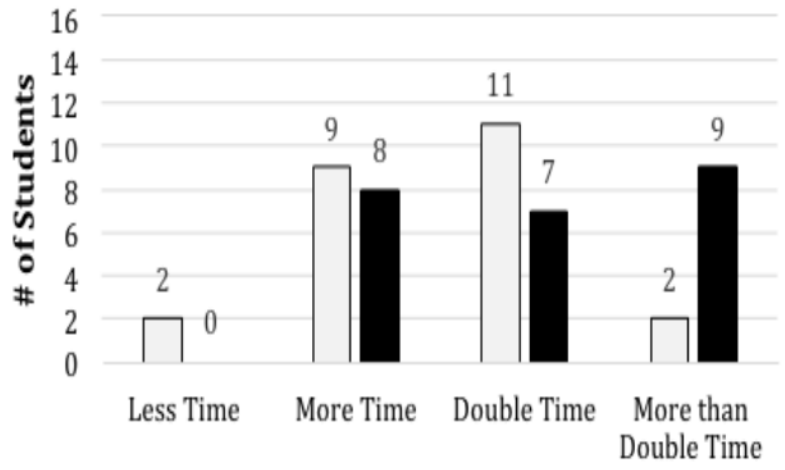


Figure 10. Task 4 Pre-Post Rubric Scores by Student.

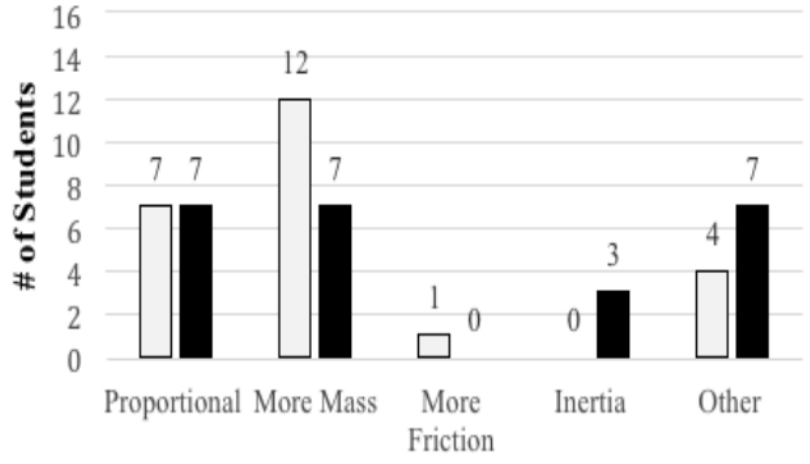
Note: Dots represent unchanged rubric scores and arrows represent pre-post changes in rubric level scores for each student. See Appendix A for rubric level definitions

The pattern of student responses provided in Task 4, displayed in Figure 11, provides further evidence of a possible progression in student understanding of inertia. On the pre-PA, the majority of students claimed that it would take more time or twice the amount of time to push two boxes, explaining that this was either because the figure would simply be pushing more mass or because the time required to push the boxes would increase in proportion to the mass. On the pre-PA, only two students correctly predicted that pushing two boxes would take more than twice the time required to push one box. On the post-PA, students were nearly evenly split among predicting that pushing two boxes would require more than twice the amount of time, more time, or twice the amount of time. Although only three students provided explanations indicating their understanding of inertia on the pre-PA administration, the majority of students invoked inertia following SLIDER instruction, with six students independently using inertia to explain the phenomena and ten students doing so after prompting ("In your class, you learned about inertia. What can you tell me about inertia that might explain why this happened?").

Prediction (Before simulation video):
How long do you predict it will take the figure to push 2 boxes?



Prediction Explanation:
Why do you think it will take ___ seconds?



Explanation (After Simulation Video):
Why do you think that happened?

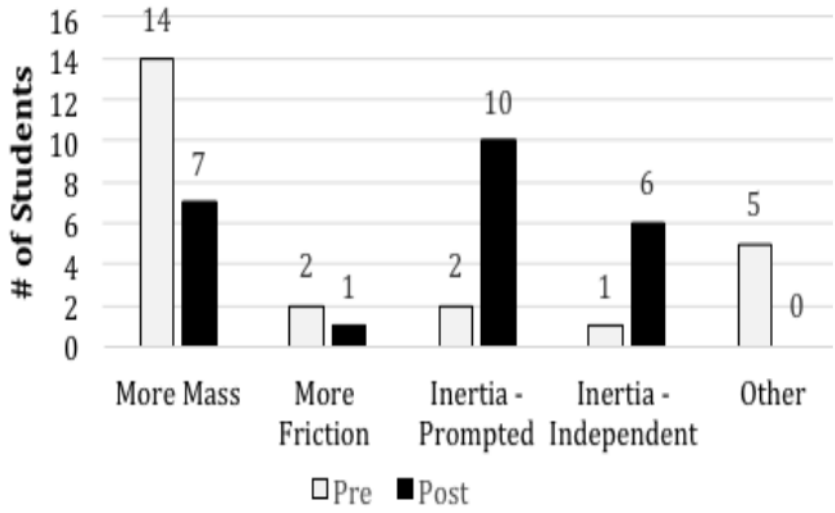


Figure 11. Task 4 Student Predictions and Explanations

ISSN:2149-8504 (online)

Table 3 provides an example of a student who provided a Level 1 response on the Pre-PA but earned a Level 3 score on the post-PA by spontaneously applying the concept of inertia both in his prediction and in his explanation of the simulation video.

Table 3.

Task 4 Illustrative Example

Pre	Post
R: When the figure was pushing one box, it took 8 seconds. Now there are two boxes. How many seconds do you think it will take for the boxes reach a speed of 70?	R: When the figure was pushing one box, it took 8 seconds. Now there are two boxes. How many seconds do you think it will take for the boxes reach a speed of 70?
S: 16 seconds.	S: (pause). 18.
R: Why do you predict ____16_____ seconds?	R: Why do you predict 18 seconds?
S: Because there are forces going the other way. So it's going to be harder to push it.	S: Because it's more than twice as much as the first one because I think it will take longer because its more…because it's harder to push something with more mass because the inertia is more, so you need more force.
R: (After Video) Why do you think this happened?	R: (After Video) Why do you think this happened?
S: Because…I don't know…because the force was greater than with one box. So with two boxes, it was greater force keeping…and you're not changing the force of the push. So if you want it to be faster, you'd have to increase the force of the push.	S: Because there is more mass, which leads to more inertia with the boxes the second time around and you need more force to push something with more inertia.
R: So it took longer because the force of the push wasn't enough?	[Scored at Rubric Level 3]
S: Yeah.	
R: Have you ever heard of inertia?	
S: No.	
[Scored at Rubric Level 1]	

Discussion

This study illustrates the potential of simulation-based PA as a method for exploring students' developing conceptions of force and motion. In their discussions of each of the four simulation-based PA tasks, students revealed the extent to which they held accurate conceptions of the force and motion concepts within the SLIDER curriculum. Implications of findings for each of the four simulation-based PA tasks are discussed below.

Task 1 was intended to be a relatively simple task used, in part, to help students become acclimated to the PA format and ease any apprehensions students may have about participating in the performance assessment interview. As expected, students found Task 1 to be simple. By the third tug-of-war scenario, all students were able to correctly determine whether there was a net force. While this result highlights the educative potential of simulation-based PAs, it also illustrates one of the complications of using PAs to measure changes in student understanding. As is the case with any assessment of pre-post learning, to the extent that the assessment

itself enables students to deepen their understanding of a concept or provides feedback that enables students to provide increasingly correct answers over the course of task administration, researchers may be limited in drawing conclusions about the degree to which results indicate pre-post differences. This difficulty is compounded when performance tasks are designed to elicit simple responses rather than, as in Tasks 2-4, eliciting students' explanations of phenomena.

Task 2 asked students to reason about the net force within the context of a motion event - a box being pushed by a figure and eventually coming to a stop after the figure has stopped pushing the box. Again, students demonstrated more sophisticated understanding at post-PA than at the pre-PA administration. Following their experience with the SLIDER curriculum, all but five students were able to correctly identify the direction of the net force when the box was in motion (being pushed and slowing down) and all students correctly answered that the box at rest had a net force of zero. The explanations students provided also became more sophisticated, with students frequently discussing the balance of applied and frictional forces within the scenario.

In Task 3, students were told that the figure pushing a box wanted to maintain a constant speed, after which they were asked, "what could the figure do to make that happen?" As the SLIDER curriculum does not include activities that explicitly ask students to reason about balanced forces in this way, this task is an example of a proximal assessment (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2001) that taps the relevant force and motion concepts but is not closely aligned to the curriculum. A greater number of students independently gave correct responses to this prompt after SLIDER instruction; however, this task remained relatively difficult, with ten students giving incorrect responses on the post-PA. Six of these students explicitly stated the alternative conception that if the figure pushed with an applied force equal to the frictional force the box would stop moving, a result that is consistent with previous conceptual development research documenting students' alternative understandings related to force and motion (McCloskey, 1983; Ioannides & 2001). Interestingly, this alternative conception appeared more commonly on the post-PA than on the pre-PA, where only two students responded that the box would stop if forces were balanced. This result may provide further evidence of the durability of this particular alternative conception and raises questions about whether and how the curriculum influences students' alternative conceptions in this area.

Task 4 represents another proximal assessment of students' developing understanding of physical science concepts. Within the SLIDER curriculum, students learn that inertia is an object's resistance to change in motion and they see a demonstration in which they make predictions and observations about the inertia of a stationary object (a dumpster being hit by a truck), but students are not asked to reason about inertia under different conditions as they are in Task 4 (i.e. one box vs. two boxes). Although this treatment of inertia within the curriculum is relatively brief, on the post-PA, the majority of students (n=16) explained the phenomena they observed in the Task 4 simulation video (i.e. dramatically increased time for the figure to push two boxes) by invoking inertia, with six students doing so spontaneously without prompting.

The results presented here lend support to the view that when it comes to revealing student understanding of difficult science concepts, simulation-based PAs may provide additional insight beyond what is obtained using traditional multiple-choice assessments, and more traditional PAs that do not involve interaction and discourse. As described above, there are a number of nuances we were able to discern through the analysis of students' responses that would not likely be evident through more traditional modes of assessment. For instance, by examining the discourse between student and researcher, we could distinguish students who spontaneously gave scientifically accurate responses from those who arrived at correct responses after engaging in further discussion with the researcher. Additionally, the study illustrates the particular benefits of simulation-based performance assessment, including the ability to simulate phenomena that would be difficult if not impossible to consistently present using physical materials. Although the time and resources invested in the development of simulation-based performance assessment tasks was considerable and may not be practical or appropriate for all assessment contexts, this approach holds promise for researchers and educators interested in gaining deeper understanding of student understanding of science concepts.

ISSN:2149-8504 (online)

These advantages notwithstanding, the study is not without its limitations. While efforts were made to select a sample representative of SLIDER students in the participating school, these results do not necessarily reflect the learning outcomes of all students who participated in the curriculum. A second limitation is the possibility of a test-retest bias. Given that the PA tasks and interview experience were likely quite novel, it is possible that students' pre-PA experience may have influenced performance on the post-PA. However, with the post-PA scheduled nearly three months following the pre-PA, we believe it is unlikely that students' remembered specific details or questions within the tasks. Additionally, with the exception of Task 1 where students watched videos illustrating the outcomes of the tug-of-war scenarios, our protocol intentionally did not provide students with "correct" answers to the PA task questions. Although the researcher who conducted the performance assessment interviews was present in the classroom prior to the pre-PA, she had spent much more time in the classroom conducting observations and focus groups with the participating students prior to the post-PA, so it is possible that students were more comfortable speaking with the researcher during their second PA experience.

Results from this study suggest a need for future research exploring innovative applications of simulation-based PA tasks. While the tasks utilized for this study required one-on-one interviews, one can envision similar tasks that could be administered online, perhaps for use by classroom teachers. Developing online simulation-based performance assessments that adequately probe student responses to generate useful assessment data presents a difficult but perhaps worthy challenge. Additionally, simulation-based PAs used in pre-post designs could be further developed by adding metacognitive items at post-PA in which students are presented with their previous responses and asked to reflect on changes in their understanding.

Conclusion

As performance assessment has emerged as a priority within the science education community, studies reporting on the administration and results of PAs will be essential. In addition to providing evidence of science learning outcomes of the SLIDER curriculum, this study illustrates the use of simulation-based PA as a promising method for gaining insight into student understanding of physical science concepts prior to and following curriculum implementation. As such, this work provides an opportunity to consider the advantages of PA over traditional modes of assessment. Similarly, this line of research raises important questions about the practical and methodological limitations of simulation-based performance assessment.

References

Alonzo, A. & Steedle, J. (2008). Developing and assessing a force and motion learning progression. Science Education, 93, 389-421.

American Association for the Advancement of Science [AAAS] Project 2061 (n.d.) [Pilot and field test data collected between 2006 and 2010]. Unpublished raw data.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME). (2014). Standards for educational and psychological testing. Washington, DC: Author

Corder, G. W., & Foreman, D. I. (2014). Nonparametric statistics: A step-by-step approach. John Wiley & Sons.

Gale, J., Wind, S., Koval, J., Dagosta, J., Ryan, M., & Usselman, M. (2016). Simulation-based performance assessment: an innovative approach to exploring understanding of physical science concepts. International Journal of Science Education, 38(14), 2284-2302.

Engelhard, G., Jr. (2013). Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. New York: Routledge.

Goetz, J. P. & LeCompte, M. D. (1984). Ethnography and qualitative design in educational research. New York: Academic Press.

Hickey, D. T., & Zuiker, S. J. (2012). Multilevel assessment for discourse, understanding, and achievement. Journal of the Learning Sciences, 21(4), 522-582.

Ioannides, C., & Vosniadou, S. (2001). The changing meanings of force: From coherence to fragmentation. Cognitive Science Quarterly, 2(1), 5–62.

Lane, S., & Stone, C. (2006). Performance assessment. In R. Brennan (Ed.), Educational Measurement, Fourth Edition (pp. 387-431). Westport, CT: American Council on Education and Praeger.

McCloskey, M. (1983). Intuitive physics. Scientific American, 248(4), 114–122.

Mislevy, R. J., Steinberg, L. S., Almond, R. G., Haertel, G. D., & Penuel, W. R. (2001). Leverage Points for Improving Educational Assessment. CSE Technical Report.

National Research Council (NRC). (2014). Developing Assessments for the Next Generation Science Standards. Committee on Developing Assessments of Science Proficiency in K-12. Board on Testing and Assessment and Board on Science Education, J.W. Pellegrino, M.R. Wilson, J.A. Koenig, and A.S. Beatty (Eds.). Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

O'Neil Jr, H. F., & Klein, D. C. D. (1997). Feasibility of machine scoring of concept maps. Los Angeles: University of California. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). Knowing what students know: The science and design of educational assessment. National Academies Press.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. Journal of Mixed Methods Research, 1(1), 77-100.

Thompson Tutwiler, Metcalf, Kamarainen, Grotzer, & Dede, (2016). A blended assessment strategy for EcoXPT: An experimentation-driven ecosystems science-based multiuser virtual environment. Paper presented at the Annual Conference of the American Educational Research Association, Washington D.C.

Usselman, M., & Ryan, M. (2014). SLIDER: Science learning integrating design, engineering and robotics. In Sneider, C. (Ed). The Go-To Guide for Engineering Curricula, Grades 6-8: Choosing and Using the Best Instructional Materials for Your Students. Corwin Press.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. Cognitive Science, 18, 123–183.

White, B. Y., & Frederiksen, J. R. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. Inquiring into inquiry learning and teaching in science, 331-370.

ISSN:2149-8504 (online)

Appendix: Simulation-Based Performance Task Rubric

Task	Incorrect		Correct	
	Level 0	Level 1	Level 2	Level 3
Task 2	Student responds that they do not know and/ or gives non-sensical responses.	Student incorrectly indicates whether there is a net force and/or the direction of the force (for any time point).	For every time point, student correctly indicates whether there is a net force and selects the correct arrows to represent the net force. (Yes,R; Yes, L: No, -)	For every time point, students correctly indicate whether there is a net force and select the correct arrows to represent the net force. (Yes,R; Yes, L: No, -) AND student compares applied vs. friction force or discusses balanced forces for any time point.
Task 3	Student responds that they do not know and/ or gives non-sensical responses.	Student responds that to maintain speed, the figure should apply a force other than 125N. Task scored as Level 1 and recommendation as one of the following: Stay at 250N Force Between Greater than 250N	Student responds that the figure should apply 125N of force but does NOT refer to balanced forces in explanation. Explanation Codes: Stopping: Student states that if applied force =125N box will stop. Prompted: Student begins with Level 1 response but through questioning arrives at Level 2 or Level 3 response. Independent: Student independently states that figure should apply 125 N of force so the forces would be balanced.	Student responds that the figure should apply 125N of force so the forces are balanced.
Task 4	Student responds that they do not know and/or gives non-sensical responses.	Student provides explanation of increased time that indicates alternative understanding of science concepts (force, motion, inertia, gravity, etc.)	Student explanation of increased time indicates accurate understanding of force and motion concepts but does not include inertia.	Student explanation of increased time indicates accurate understanding of inertia.