Exploring Student Understanding of Force and Motion

Using a Simulation-Based Performance Assessment

Jessica Gale, Jayma Koval, Stefanie Wind, Joe Dagosta, Mike Ryan, Marion Usselman

Georgia Institute of Technology

Center for Education Integrating Science, Mathematics, and Computing

Paper presented at 2016 Annual Meeting

ofthe American Educational Research Association, Washington, D.C.

Abstract

Performance assessment (PA) has been increasingly advocated as a method for measuring students' conceptual understanding of scientific phenomena. This paper illustrates the use of a simulation-based PA to measure 8th grade students' understanding of physical science concepts taught via the experimental SLIDER (Science Learning Integrating Design, Engineering, and Robotics) curriculum. The performance assessment consisted of 4 tasks that use simulation videos to illustrate key force and motion concepts (net force, acceleration, inertia). These tasks were administered to a stratified sample of 24 students in one school prior to and following implementation of the SLIDER curricula. The patterns of student responses revealed through qualitative analysis provide preliminary evidence of student learning over the course of the curriculum implementation period. Limitations and implications for future research are discussed.

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). Within science education, PAs have been touted as essential indicators of student mastery of 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 also articulated by the Committee on Developing Assessments of Science Proficiency in K-12 for the Next-Generation Science Standards (NGSS), with the 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). PA methods used in science education have included tasks that ask 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, resources, and costs. Given these limitations, there are few examples of research utilizing performance assessments to measure developments in science students’ conceptual understanding over the course of curricular interventions. Simulation-based assessments offer a potential compromise. The

Standards note these assessment formats may be especially appropriate in contexts where "actual task performance might be costly or dangerous" (AERA, APA, & NCME, 2014, p. 78). The purpose of this paper is to illustrate the use of simulation-based performance assessment (PA) within the context of a design-based implementation research (DBIR) project. This paper describes simulation-based PA tasks designed to assess 8th grade students' understanding of physical science concepts. The paper also provides preliminary data illustrating changes in conceptual understanding in a sample (N=24) of students in one 8th grade classroom.

Methods

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

Science Learning Integrating Design, Engineering, and Robotics (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-achievement suburban classrooms to relatively low-proficiency, low-income rural schools. 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 (increased accidents at a dangerous intersection). 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 slider.gatech.edu.

Participants

The PA was administered to 24 eighth grade physical science students attending a middle school in an affluent suburban community during the 2014-15 school year. All of the students were taught by one teacher implementing the SLIDER curriculum during that 2014-15 school year. The authors selected students from this particular teacher's classes because the teacher exhibited the highest fidelity of implementation of the SLIDER 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 measures of student achievement, students were classified into three groups based on their achievement levels on the SLIDER Unit 2 pre-assessment (high, medium, and low). The second stage of the student selection 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 select 24 students (eight students from each achievement level) who had consistent attendance and had actively participated in previous SLIDER activities. The teacher was not informed that the three columns in the matrix represented student grouping based on achievement.

Instrument – Performance Assessment Tasks

The PA instrument includes four tasks, developed in collaboration with the SLIDER curriculum team to assess student understanding of major concepts addressed in the curriculum:

net force, acceleration, friction, balanced forces, and inertia. The tasks were developed by adapting simulations from the University of Boulder PhET Interactive Simulation website (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 are described below.

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 on the rope. The task proceeded with three scenarios in which 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 will be a net force, they were shown two arrows, a large arrow and a small arrow, and asked to choose an 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

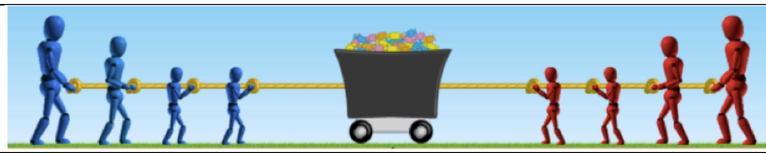
Scenario 1	
4 vs. 4	
Scenario 2	
4 vs. 1	
Scenario 3	
3 vs. 4	

Figure 1. Task 1: Net Force

Task 2: An Object in Motion Task 2, depicted in Figure 2 below, assessed students' understanding of net force using a simulation in which a figure pushes a box along a surface that 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 place an arrow on the illustration 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").

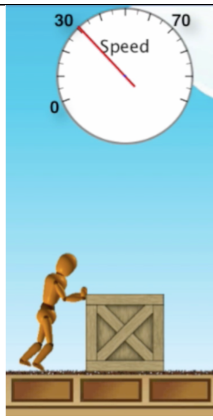
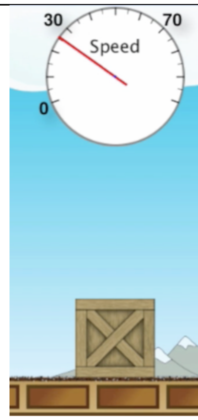
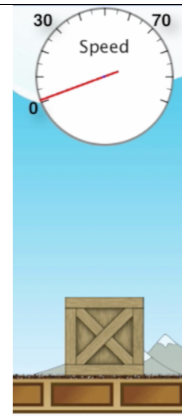
Time-point 1	Time-point 2	Time-point 3
		
Speed of the box is increasing.	Figure has fallen away and the speed of the box is decreasing.	The box is at rest.

Figure 2. Task 2: An Object in Motion

Task 3: Balanced Forces. In Task 3, depicted in Figure 3 below, 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, in order to determine whether students held the common misconception that balancing forces would cause the object to stop, when students responded that the figure should push with more than 125N of force, the researcher asked “what do you think would happen if the figure pushed with 125N?”

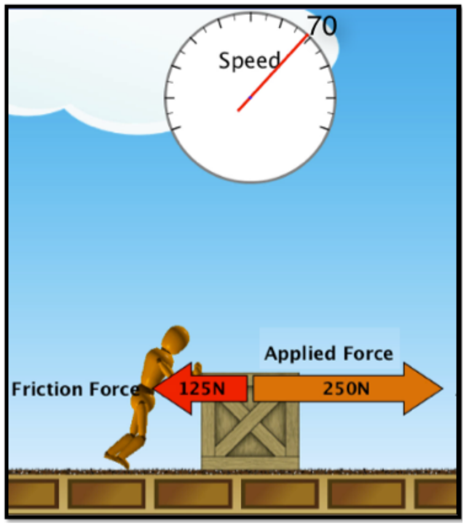


Figure 3. Task 3: Balanced Forces

Task 4: Inertia. Task 4, depicted in Figure 4 below, 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 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 2 boxes to the target speed of 70. Students were then asked to explain why it took so much longer for the figure to push 2 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?").

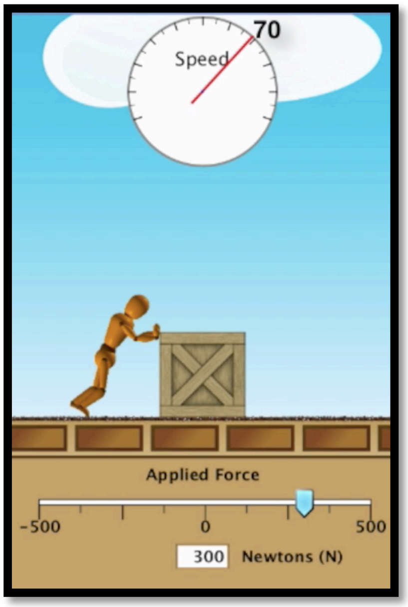
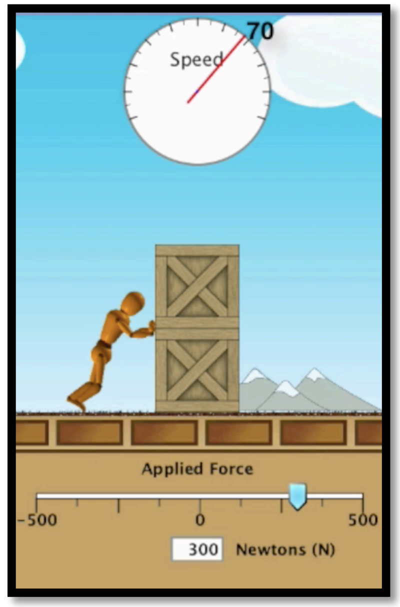
<table>
<tr><th>Scenario 1</th><th>Scenario 2</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>1 Box, 300 N of Applied Force</td><td>2 Boxes, 300 N of Applied Force</td></tr>
</table>

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 and compared for each of the twenty-four participating students. Because student responses for PA Task 1 were limited to just 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 so 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 task rubrics iteratively developed by the research team. Rubrics 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 understanding of targeted science concepts. *Explanation* codes were utilized to categorize the explanations and predictions students provided within the tasks and to indicate whether students arrived at their ultimate responses independently or through follow-up questions from the researcher (e.g. “coaching”). Task rubrics (Appendix A) 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 over 90% agreement between coders for each tasks. Any remaining coding discrepancies were resolved through discussion between the two coders.

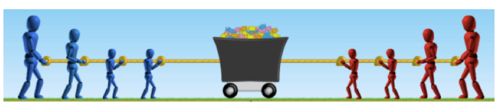
Results

This section presents preliminary results and illustrative examples for each PA task.

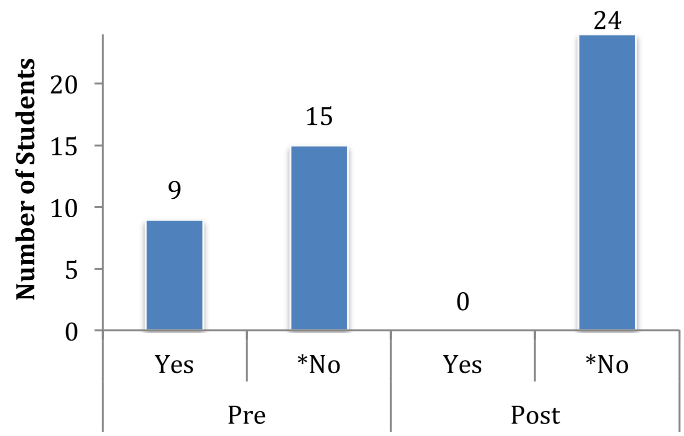
Task 1

Figure 5 below illustrates the frequency of student responses to the Task 1 prompts for the three tug-of-war scenarios.

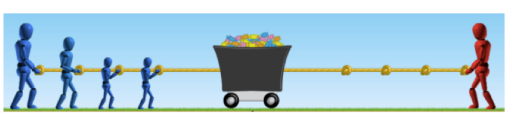
Scenario 1



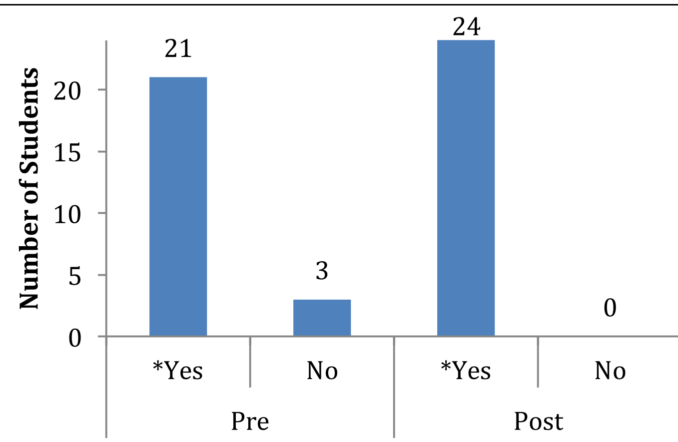
"Is there a net force?"



Scenario 2

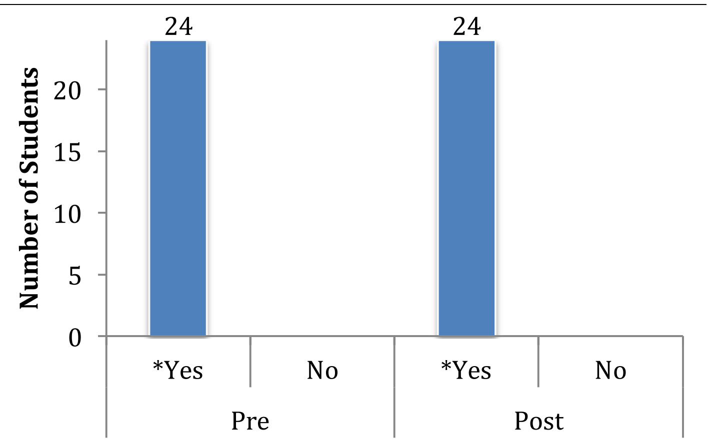


"Is there a net force?"



Scenario 3

“Is there a net force?”



Note: Correct responses indicated by a (*).

Figure 5. Student Responses for Task 1 Question: Is there a net force?

Student responses to the Task 1 prompt, “Is there a net force?” suggest potential differences between pre-and post response patterns. As indicated in Figure 1, nine students answered incorrectly for the relatively simple scenario 1 and few or no students answered incorrectly at either pre- or post-PA for the more difficult scenarios two and three. Although this pattern may suggest that students who began the task with a lack of understanding of net force learned the basic concept over the course of the simulation-based PA, 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 meaning of “net force” rather than developing an accurate understanding of the concept. 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 prior to the presentation of much more conceptually difficult tasks that would require students to provide explanations of force and motion phenomena depicted in simulations. The ease with which students correctly responded to the prompts suggests that Task 1 was successful in this regard.

Task 2

Recall that in Task 2, students watched a video 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. Students' overall task performance, as evaluated using a holistic coding rubric, is presented in figure 6 below. Figure 7 below 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 the purpose of analysis.

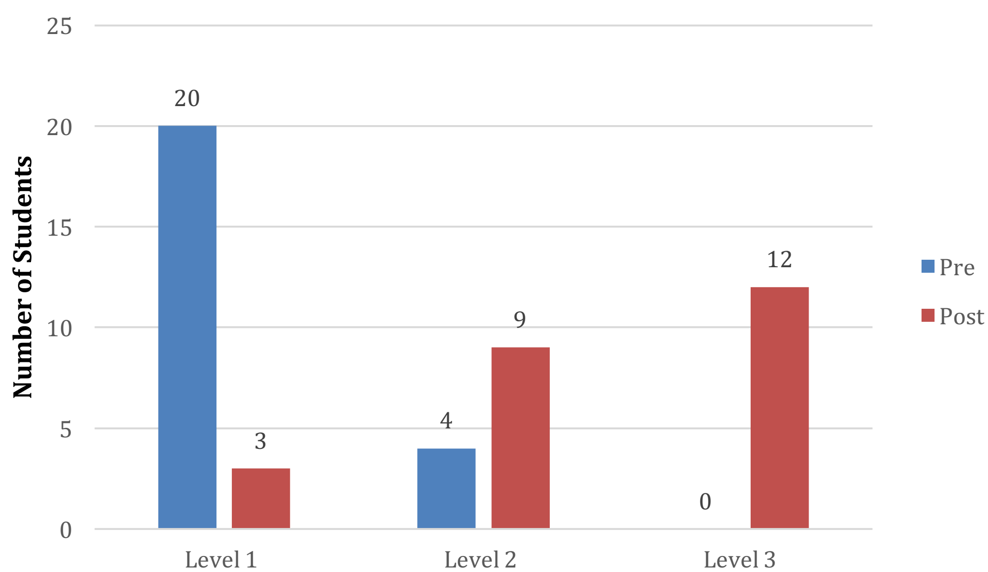
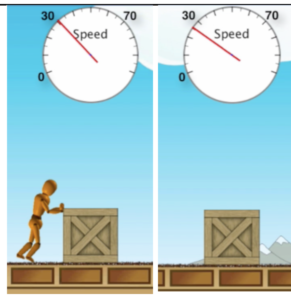
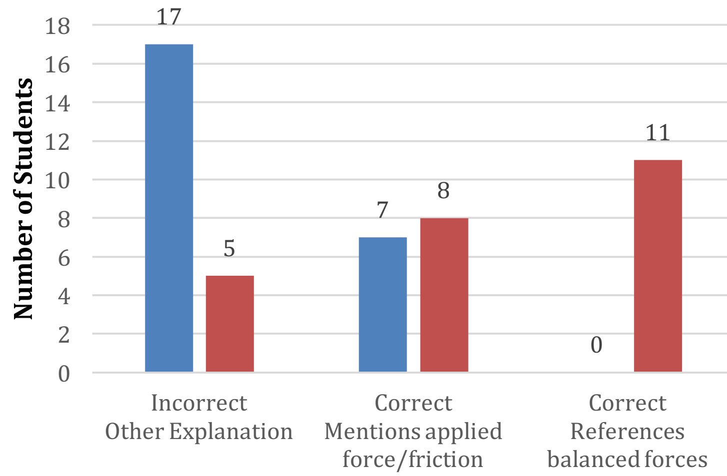


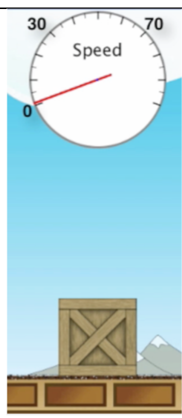
Figure 6. Task 2 Holistic Coding Results. See Appendix A for rubric level definitions.

Box in Motion





Box at Rest



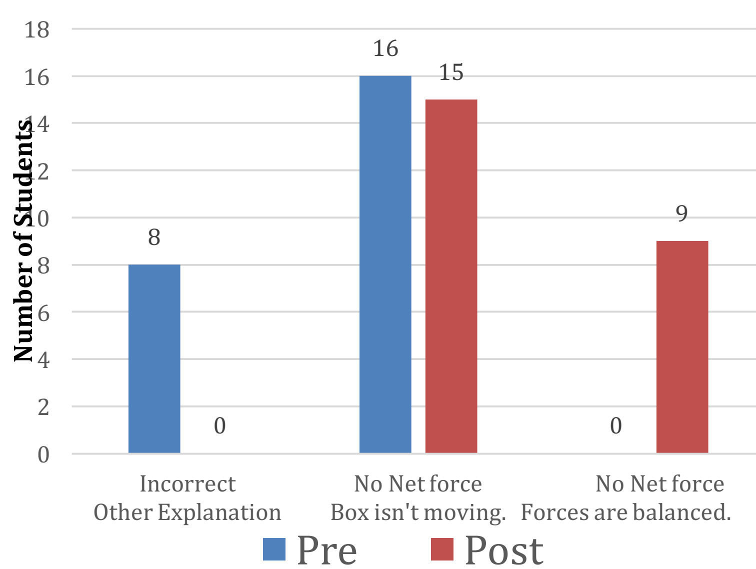


Figure 7. Task 2 Student Explanations

Taken together, student responses coded using the holistic and explanation rubrics illustrate a clear shift in student understanding of the targeted physical science concepts assessed by Task 2. Prior to SLIDER implementation, 20 students gave a Level 1 response, incorrectly stating whether there was a net force and/or indicating the wrong direction of the net force and relatively few students provided explanations that referred to applied force and/or friction (Level 2) or compared applied and frictional forces (Level 3). In contrast, following SLIDER, the vast majority of students were able to provide responses that earned a Level 2 or Level 3 score on the holistic rubric by correctly indicating whether there was a net force and correctly placing the arrow on the illustration to indicate the direction of the net force at each time-point. Similarly, at the post-administration, students were far more likely to explain their responses by comparing applied vs. frictional forces or explicitly discussing balanced forces. This shift is evident in the illustrative example presented in Table 1 below, in which the student provides a Level 1 response prior to SLIDER and a Level 3 response following curriculum implementation.

Table 1

Task 2 Illustrative Pre-Post Example

Pre-SLIDER	Post-SLIDER
Timepoint 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. Timepoint 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. Timepoint 3:	Timepoint 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. Timepoint 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: 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. [Response scored at Rubric Level 1]	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. Timepoint 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 box can't move forward because there is no one to push it forward. [Response scored at Rubric Level 3]
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 below illustrates the distribution of students' scores on the holistic coding rubric for Task 3. These results suggest some development in students' understanding of how balanced forces operate when an object is in motion between the pre- and post- administration of the PA, 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 suggests that this was a particularly difficult task for many students.

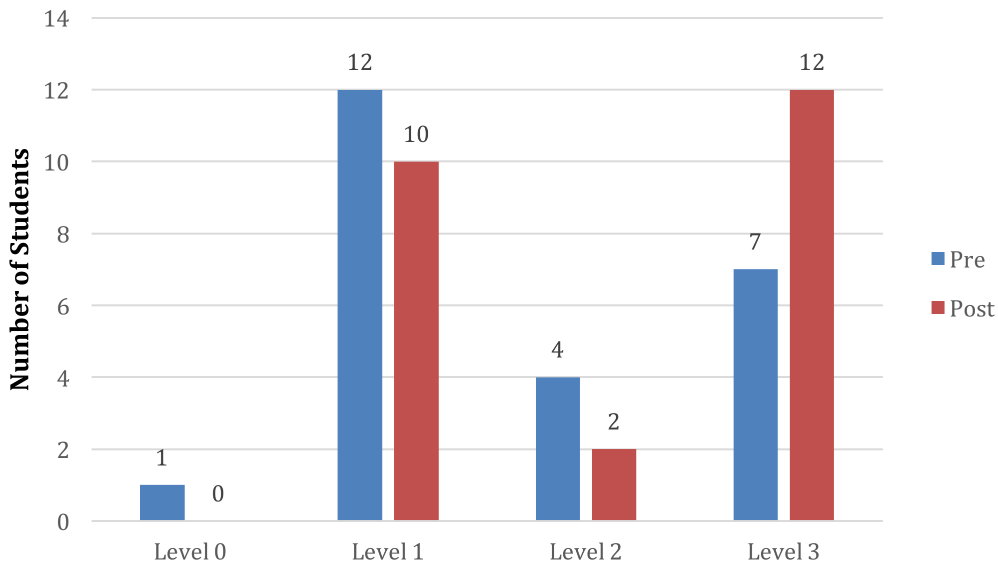


Figure 8. Task 3 Holistic Coding Results. See Appendix A for rubric level definitions.

Figure 9 below 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 (2 at pre-PA and 6 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 coaching 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, which we refer to as “coaching”, was not necessarily to lead students to change their answers, 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 coaching and that students were somewhat more likely to independently provide correct responses following their experience with the SLIDER curriculum.

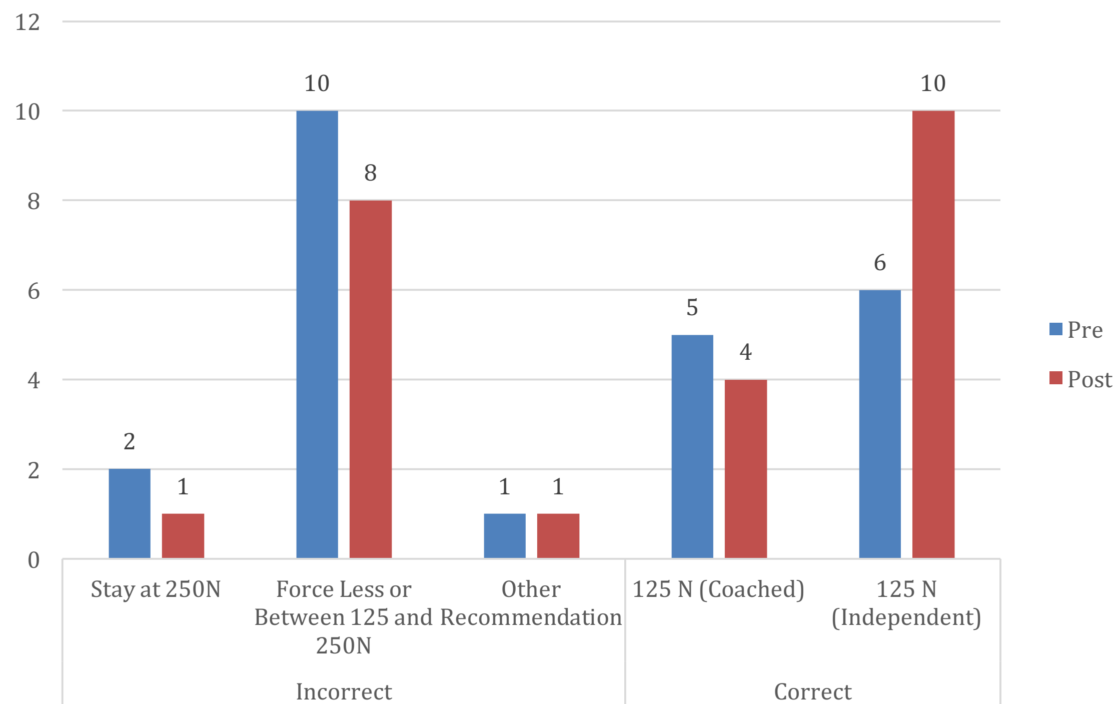


Figure 9. Student Responses to Task 3 Question: What could the figure do to keep speed at 70?

Table 2 below presents an illustrative example of one students’ responses for pre- and post-SLIDER administrations of Task 3. Prior to the SLIDER curriculum, this 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. In the post-administration of the task, the student seems to have revised his understanding to confirm his initial conception and independently responds that balanced forces would produce a constant speed.

Table 2

Task 3 Illustrative Pre-Post Example

Pre-SLIDER	Post-SLIDER
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."	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.
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 to reach a certain speed when pushing two boxes versus one box. Figure 10 below 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 were able to explicitly apply the concept of inertia to explain what they observed in the simulation.

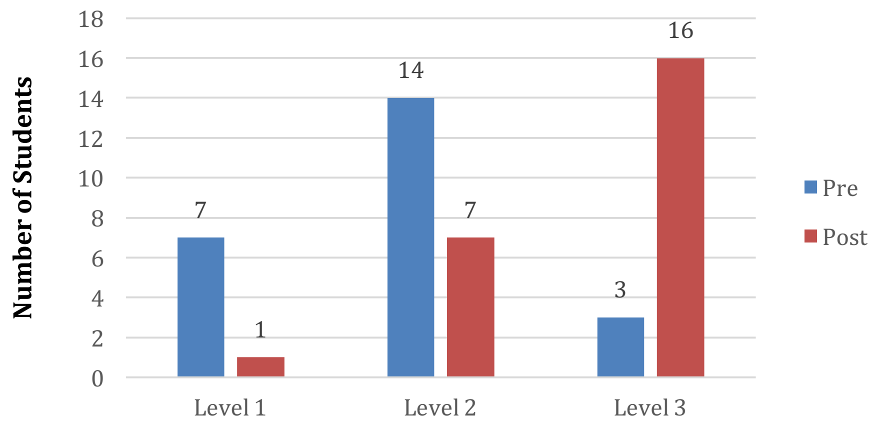


Figure 10. Task 4 Holistic Coding Results. See appendix A for rubric level definitions.

The distribution of student responses provided in Task 4 provides further evidence of a possible progression in student understanding of inertia. Figure 11 below displays students' predictions and explanations for Task 4.

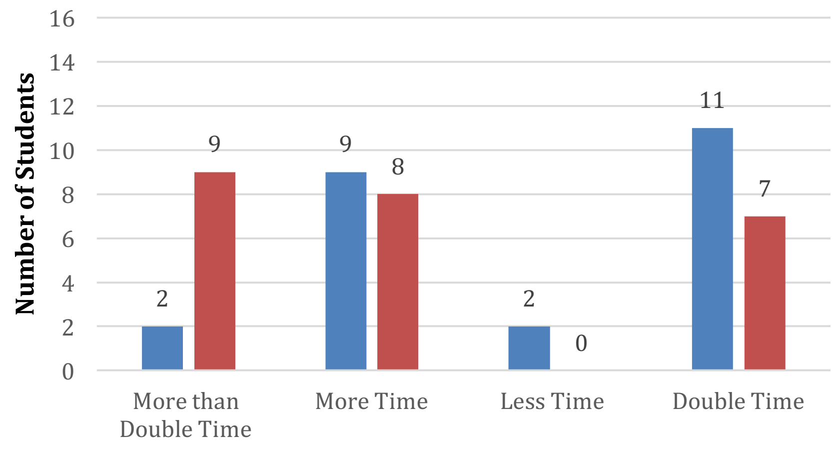
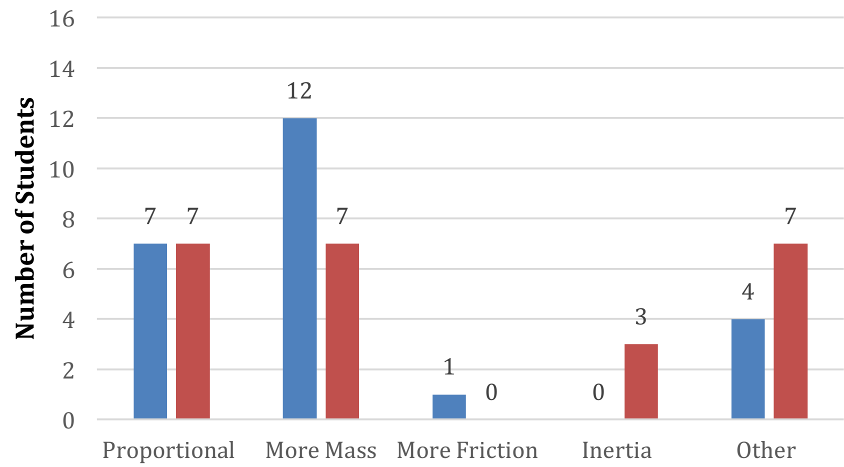
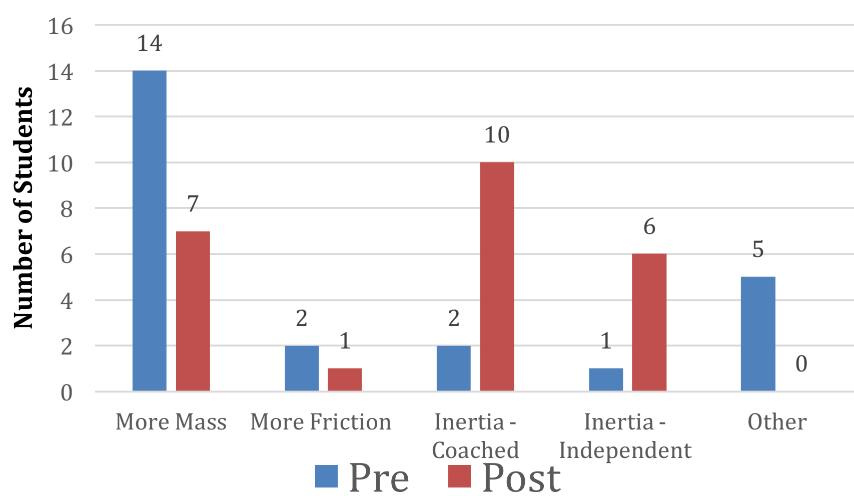
<table>
<tr>
<td>Prediction (Before watching simulation video): How long do you predict it will take the figure to push 2 boxes?</td>
<td>
</td>
</tr>
<tr>
<td>Prediction Explanation: Why do you think it will take ___ seconds?</td>
<td>
</td>
</tr>
<tr>
<td>Explanation (After Simulation Video): Why do you think that happened?</td>
<td>
</td>
</tr>
</table>

Figure 11. Task 4 Student Predictions and Explanations

On the pre-PA, the majority of students thought it would take more time or twice the amount of time to push two boxes, explaining that this was because the figure would be pushing more mass or that the time required to push the boxes would increase in proportion to the increased 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 split among predicting that pushing two boxes would require more than twice the amount of time, more time, or twice the amount of time. While only three students provided explanations indicating their understanding of inertia on the pre-PA administration, the majority of students provided explanations invoking inertia following SLIDER instruction, with six students independently using inertia to explain the phenomena and ten students doing so after being prompted ("In your class, you learned about inertia. What can you tell me about inertia that might explain why this happened?"). Table 3 below 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.

Task 4 Illustrative Pre-Post Example

Pre-SLIDER	Post-SLIDER
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. R: Why do you predict ____16_____ seconds?	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: (pause). 18. R: Why do you predict 18 seconds? S: Because it's more than twice as much as

S: Because there are forces going the other way. So it’s going to be harder to push it. R: (After Video) Why do you think this happened? S: Because—I don’t know. Because the force was greater than the box one. 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. R: So it took longer because the force of the push wasn’t enough? S: Yeah. R: Have you ever heard of inertia? S: No.	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? 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.
S = Student R= Researcher	

Discussion

This case study provides preliminary evidence of student learning within the context of the SLIDER project. Results of each of the four PA tasks are described 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. 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 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 degree to which the applied and frictional forces within the scenario were balanced.

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 for many students, with ten students giving incorrect responses on the post-PA. Six of these students explicitly stated their alternative conception that if the figure pushed with an applied force equal to the frictional force, the box would stop moving. 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. Future work will explore the extent to which student pre- and post- responses on Task 3 align with proposed learning progressions for force and motion concepts (e.g. Alonzo & Steedle, 2008).

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. 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 2 boxes) by invoking inertia, with six students doing so spontaneously without prompting from the researcher.

The data presented here lend support to the view that when it comes to revealing student understanding of difficult science concepts, performance assessments may have advantages over traditional multiple-choice assessments. As described above, there are a number of nuances we were able to discern through the qualitative analysis of students' performance assessment responses that would not likely be evident through more traditional modes of assessment. For instance, by examining the discourse between students and researcher, we could distinguish between students who spontaneously gave scientifically accurate responses from those where students' arrived at correct responses after being prompted. 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 to students using physical materials. Although the time and resources invested in the development of simulation-based performance assessment tasks was considerable, we believe that this approach holds promise for researchers and educators interested in gaining deeper understanding of student understanding of science concepts.

These advantages notwithstanding, the study is not without its limitations. While efforts were made to select a sample representative of students who completed the SLIDER curriculum 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 tasks. 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 study 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. Finally, the summary analyses presented here do not track the development of conceptual understanding at the student level. Without further analyses we cannot, for example, report the proportion of students who maintained, progressed, or regressed with regard to their understanding of the targeted physical science concepts from pre- to post-PA. Future work will include additional analyses that enable us to better discern these student-level response patterns.

Another avenue of future research would involve the continuous refinement of simulation-based performance assessment tasks. While the tasks utilized for this study required one-on-one interviews with students, one can envision developing similar tasks that could be administered online, perhaps for use by classroom teachers. Developing an online simulation-based performance assessment that adequately probes student responses and generates useful assessment data presents a difficult but perhaps worthy challenge. Additionally, researcher-administered 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 conceptual understanding over the course of curriculum implementation.

Conclusion

As performance assessment has emerged as a priority within the science education community, studies that report on the administration and results of PAs will be essential. In addition to providing evidence of science learning following implementation of the SLIDER curriculum, this study illustrates the use of simulation-based performance assessment to gain 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 performance assessment.

References

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

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). *Ethnogaphy and qualitative design in educational research.* New York: Academic Press.

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.

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.

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

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

Task 2 Holistic Codes			
Incorrect		**Correct**	
Level 0	**Level 1**	**Level 2**	**Level 3**
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 2 Event Codes				
Event	Incorrect		Correct	
	Level 0	Level 1	Level 2	Level 3
Box is Speeding Up/Slowing Down (T1/T2)	Student responds that they do not know and/or gives non-sensical responses.	Student gives other explanation of net force.	Explanation mentions applied force and/or friction without comparison of forces or reference to unbalanced forces. (EX: For T1 - Student attributes direction of net force arrow to the figure pushing the box or For T2 – Student attributes direction of force to box slowing down.)	Explanation compares applied force vs. friction (For T1 applied > friction, for T2 friction>applied,) or refers to unbalanced forces.
Box is at Rest (T3)	Student responds that they do not know and/or gives non-sensical responses.	Student gives other (incorrect) explanation of net force.	For T3 – Student states that because the box isn't moving, there is no net force.	For T3, student indicates no net force because the forces are balanced.

Task 3 Holistic Codes			
Incorrect		**Correct**	
Level 0	**Level 1**	**Level 2**	**Level 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. Code entire task as Level 1 and recommendation as one of the following: **Stay at 250N** **Force Between** **Greater than 250N** (ONLY code at Level 1 if student gives one of the above responses and does not change their mind after prompting).	Student responds that the figure should apply 125N of force but does NOT refer to balanced forces in explanation.	Student responds that the figure should apply 125N of force so the forces are balanced.

Task 3 Explanation Codes			
Other: Student gives other explanation indicative of alternative understanding.	**Stopping:** Student states that if applied force =125N box will stop.	**Coached:** 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.

Task 4 Holistic Codes			
Incorrect		Correct	
Level 0	**Level 1**	**Level 2**	**Level 3**
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.

	Task 4 Prediction/Explanation Codes			
	Incorrect		Correct	
	Level 0	Level 1	Level 2	Level 3
Task 4 Prediction ("How many seconds….?")	Student does not give a prediction.	**Less:** Student predicts that it will take less time to push 2 boxes than 1 box.	**Longer:** Student predicts that it will take longer to push 2 boxes to reach a speed of 70 than it took to push 1 box. **Double**: Student predicts that it will take double the time to push 2 boxes.	**More than Double:** Student predicts that it will take more than double the time.
Explanation of Prediction (pre-video) ("Why did you predict…?")	Student responds that they do not know and/or gives a non-sensical response.	**Other:** Student gives other explanation indicative of alternative understanding.	**More Mass** – Student references increased mass/weight to explain prediction. **More Friction** – Student references increased friction to explain prediction. **Proportional** – Student uses proportional reasoning (EX: if one box took 8 sec., 2 box will take 16).	**Inertia:** References inertia to explain prediction (more than double).
Explanation (post-video) ("Why do you think that happened?")	Student responds that they do not know and/or gives a non-sensical response.	**Other:** Student gives other explanation indicative of alternative understanding.	**More Mass** – Student references increased mass/weight to explain prediction. **More Friction** – Student references increased friction to explain prediction. **Proportional** – Student uses proportional reasoning (EX: if one box took 8 sec., 2 box will take 16). **Coached** – Student begins with L1 or L2 response but through questioning explain that increased time was due to inertia.	**Inertia:** Student independently references inertia to explain increased time (in video).