



Use of physics simulations in whole class and small group settings: Comparative case studies



A. Lynn Stephens*, John J. Clement

TECS Department, College of Education, University of Massachusetts, Furcolo Hall, Amherst, MA 01003, USA

ARTICLE INFO

Article history:

Received 19 July 2014

Received in revised form

25 February 2015

Accepted 28 February 2015

Available online 9 March 2015

Keywords:

Cooperative/collaborative learning

Improving classroom teaching

Pedagogical issues

Simulations

Teaching/learning strategies

ABSTRACT

This study investigates student interactions with simulations, and teacher support of those interactions, within naturalistic high school classroom settings. Two lesson sequences were conducted, one in 11 and one in 8 physics class sections, where roughly half the sections used the simulations in a small group format and matched sections used them in a whole class format. Unexpected pre/post results, previously reported, had raised questions about why whole class students, who had engaged in discussion about the simulations while observing them projected in front of the class, had performed just as well as small group students with hands-on keyboards. The present study addresses these earlier results with case studies (four matched sets of classes) of student and teacher activity during class discussions in one of the lesson sequences. Comparative analyses using classroom videotapes and student written work reveal little evidence for an advantage for the small group students for any of the conceptual and perceptual factors examined; in fact, if anything, there was a slight trend in favor of students in the whole class condition. We infer that the two formats have counter-balancing strengths and weaknesses. We recommend a mixture of the two and suggest several implications for design of instructional simulations.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The purpose of this study is to investigate student interactions with simulations, and teacher support of those interactions, within naturalistic¹ high school physics classroom settings. We ask what differences there might be between whole class and small group discussions during use of simulations. Constructivist educators have stressed the importance of learning by doing, which, in our experience, has been interpreted by many teachers to mean that students must have their own hands on the keyboards. However, we have noticed that students may misinterpret, or simply miss, important information in a simulation. Because simulations are intended to convey dynamic visual information, teachers may be tempted to believe that simulations are automatically effective in communicating complex models to students. However, research such as Lowe (2003) has shown that comprehension of animations is dependent on appropriate prior knowledge structures. We have observed interesting student reasoning and interesting teacher support moves designed to promote reasoning and comprehension during use of simulations in both whole class and small group contexts; this has led us to look more deeply at what is occurring during these discussions. Our motivating question is, does one format have strengths that the other does not? Though this is a complex question, we can begin to address it by comparing several factors at work in the class discussions.

* Corresponding author. Present address: Center for Knowledge Communication, School of Computer Science, 140 Governors Drive, University of Massachusetts, Amherst, MA 01003-9264, USA. Tel.: +1 413 545 2744, fax: +1 413 545 1249.

E-mail address: lstephen@umass.edu (A.L. Stephens).

¹ We mean 'naturalistic' as opposed to 'in the laboratory.' We tried to interfere with the classroom environment as little as possible because our interest is in what can happen in a real-world classroom.

We observed classes using simulations in one of two formats, a small group hands-on format or a whole class discussion format during which a single computer was used to project the simulation in front of the class. We will review the pre/post results from Stephens (2012) and then focus on two broad issues: the extent to which students engaged with conceptual issues and the extent to which visual features were recognized, used and supported. Comparative case study analyses of four matched sets of classes identify differences and similarities between the class sections in each matched set, as revealed in classroom videotapes and student written work.²

2. Theoretical background

Studies have investigated the effects of *instructional guidance for simulations* when guidance was provided within the learning materials or by the teachers (review by Cook, 2006; Reid, Zhang, & Chen, 2003) and have recommended such actions as providing interpretive support and minimizing cognitive load. The de Jong and van Joolingen (1998) review of simulation use in discovery learning contexts cited the importance of structuring and supporting students' work in ways to prevent difficulties. However, there do not appear to be many studies that address the question of how best to provide instructional guidance for simulations and animations in the context of whole class discussion.

Researchers have studied the use of simulations and other digital tools by *small groups* and by *individual students* (Adams et al., 2008a; Buckley, 2000; Linn, 2003; Williams, Linn, Ammon, & Gearhart, 2004; Windschitl & Andre, 1998; and Zietsman & Hewson, 1986). Among the potential advantages described for these tools are that they can increase engagement, that teachers can use them to "help students make their thinking visible," and that much of this software provides students the opportunity to customize their own modeling tools. Another potential benefit is that animated graphics can show changes over time (review by Cook, 2006) although these can also produce cognitive overload and actually hinder novice learning (Lowe, 2003; Tversky, Morrisson, & Betrancourt, 2002). Therefore, novices may need to be cued to details of motion in animated graphics (Rieber, 1990).

Hands-on activity afforded by small group work would appear to offer students a more *active learning experience* with simulations than would a whole class format. In the context of think-aloud interviews, for example, Adams et al. (2008a), indicate that simulations can be highly effective, but only if the student's interaction is directed by the student's own questioning. This kind of self-directed interaction with a simulation would seem to require individual or small group work with hands-on-keyboard opportunities. Considering this, and the fact that the teachers in our study have stated they prefer to allow students to work with simulations in small groups and feel experienced teaching in that format, it might be expected that the small group format would work better for them than would a whole class format. On the other hand, studies have reported a variety of issues concerning the effective use of small group discussions in science classes (some of which used simulations), including that students can exhibit a low level of engagement with tasks (Bennett, Hogarth, Lubben, Campbell, & Robinson, 2010), in contrast to findings cited above.

Good practices for teacher response in *whole class discussion* have been and continue to be informed by the work of Chin (2006), Hammer (1995), Hogan and Pressley (1997), and McNeill and Pimentel (2009) among others. In general, these recommend that the teacher play a role in 1) drawing out student reasoning and 2) scaffolding certain kinds of reasoning where students have difficulty. However, these studies did not focus on use of interactive simulations (but see Raghavan, Sartoris, & Glaser, 1998, for a counterexample). What kinds of teacher responses are optimal during class discussion may be affected by use of an interactive simulation that has been designed to provide feedback and to serve as an expert voice. Some believe we know very little about how to use animation effectively in instruction (Jones, Jordan, & Stillings, 2001). Principles suggested by theory and by laboratory work with students using simulations (Lowe, 2003; Mayer & Moreno, 2002) would appear to need further validation in science classroom contexts (Cook, 2006), and may well have to be modified to be usable by teachers employing available simulations with available, frequently limited, classroom hardware.

The present study is of classrooms engaged in *model-based learning* in science. Studies of expert scientists and of science students conclude that the ability to generate and evaluate mental models appears to be a crucial aspect of scientists' thinking (Clement, 2008; Darden, 1991; Nersessian, 1995) and of student thinking (Clement & Ramirez, 2008; Gentner & Gentner, 1983; Nunez-Oviedo & Clement, 2008). The pedagogical approach of the teachers in this study can best be characterized as *guided inquiry* (Bell, Smetana, & Binns, 2005; Hammer, 1995), in which students are supported by the teacher in lessons that are neither pure inquiry nor pure lecture but somewhere in between. Studies of model-based instruction (Hestenes, 1987; Krajcik, McNeill, & Reiser, 2008; and Schwarz et al., 2009) emphasize that complex scientific models can be constructed using prior knowledge ideas and reasoning resources of students, but that this usually requires scaffolding from external supports of various kinds. Minstrell and Kraus (2005), Williams and Clement (2015), and Windschitl, Thompson, Braaten, and Stroupe (2012) have recently identified many interesting strategies for teachers to use in dealing with conceptual difficulties encountered during the learning of complex models in science. Findings from social constructivism (Hogan & Pressley, 1997) have led to a belief that classroom discussion that includes teacher-student or student-student exchanges can be an important and helpful component of model-based learning; we note that these can occur in either small group or whole class contexts.

Little work has been done comparing *small group vs. whole class formats*. Wu and Huang (2007) compared a single teacher-centered and a single student-centered class using physics simulations where the classroom formats were similar to the whole class and small group formats we describe. They found no overall difference in pre/post conceptual gains between the two groups, although they found qualitative differences in cognitive and behavioral engagement. The present study takes a somewhat different tack; rather than focusing on engagement, we focus on students' ability to make use of the visuals in order to address certain difficult concepts and the extent to which discussions dealt with this. Smetana & Bell (2009) compared the use of chemistry simulations in a single small group class and a single whole class discussion and found no significant difference in pre/post gains; they suggest that future research involving more varied populations and additional teachers and classrooms is needed. The present study, along with a related study of a second lesson sequence (Stephens, 2012), aims to contribute to that goal and to investigate more deeply some of the underlying factors at work in the two modes.

² Levels of physics courses included in this study, from least to greatest difficulty: CP = College Preparatory, HP = Honors, AP = Advanced Placement.

3. Research questions

For this investigation into potential differences between whole class and small group discussions during use of simulations, we focus on the following Research Questions:

- To what extent were crucial physics concepts dealt with by teachers and students in the whole class and small group conditions?
- To what extent were key visual features of the simulation recognized and used in the two conditions?

4. Material and methods

4.1. Rationale for mixed methods approach

Our primary purpose for planning videotape and transcript analysis was to increase our understanding of teaching and learning processes. More particularly, with respect to the eight class sections discussed here, we wished to investigate possible differences between whole class and small group teaching and learning modes that were invoked to support conceptual learning during science classroom discussions. Our primary method was to conduct intensive video analysis in qualitative case studies. We wanted to supplement this with the results of the pre/post tests for these sequences. In general, pre/post test analyses could be completed quickly for all observed classes, while the more intensive videotape analyses on selected lesson sequences took much longer to complete. When designing the larger study, our plan to implement video analysis precluded large samples. This issue points up a tension we have always felt when designing such studies, the tradeoff between the need for a sample size small enough to be manageable for qualitative analysis and the need for very large numbers of classes for rigorous quantitative analysis. Our solution in designing this series of studies (along with a companion set, Stephens, 2012) was to focus on the qualitative analysis with a manageable sample size and to use the results of quantitative analysis for unusually narrow purposes. The primary purpose was not to attempt to project the quantitative results to a larger population outside the study. Rather, for each of the studies discussed here, the results of the pre/post analysis yielded quantitative information about the sample inside the study, and served to motivate, inform, and constrain the design of the main qualitative study of the sample. Thus we decided on a rather narrow use of quantitative measures as part of a mixed methods design.

4.2. Participants and setting

The analyses to be described here involve class sections from two physics teachers at a high school in a suburban college town. The teachers were purposefully selected; they had to be willing to teach model-based lessons and to foster discussions in both whole class and small group settings, and they had to be willing and able to use computer simulations as part of their lesson plans. Class sections taught by a given teacher were purposefully selected for analysis according to whether they fit the following criteria, in which case they were considered to form matched sets of classes. The teacher must have been teaching at least two comparable sections in a given semester and conducted the lesson sequence in at least one section in a whole class format and in at least one other section in a small group format. Teachers' evaluations and records were relied upon to determine that the sections within a set had students comparable in terms of age and demonstrated levels of aptitude for the content of the course as evidenced by their prior work in the course. In addition, the class sections in each set must have been provided similar levels of preparedness for the lesson, as indicated by the teachers' records of their lesson plans. Finally, the lesson sequence as taught in the two formats must have been similar (see Materials and Procedure below) and the class sections must have been allowed similar amounts of time on the lessons and the pre- and post-tests. Fifteen lesson sequences were observed; seven sequences (and one teacher) were dropped from analysis because they did not meet the above criteria, leaving eight sequences from two teachers to be subjected to analysis.

Once it was determined that class sections were matched, they were assigned to the whole class (WC) or small group (SG) condition for the lesson sequence according to practical logistical considerations, such as how much time there would be before and after the class to rearrange equipment. Class sections within each matched set met in the same rooms. One or both authors observed all lesson sequences. The authors conducted follow-up interviews with the teachers.

These eight class sections comprised four matched sets as indicated in Table 1, $N = 150$. Teacher A taught this as a two-day sequence while Teacher B taught it as a one-day lesson. Therefore, 10 videotapes were collected for this lesson sequence from the eight class sections. The intention is not to draw comparisons between different teachers but to compare each teacher's small group lesson to the same teacher's whole class lesson of the same matched set.

4.3. Materials and procedure

Although materials varied slightly for each level of physics, within each matched set (as described above), the teacher used identical materials in the two conditions but varied the way in which the simulation was used. In the whole class condition, each teacher used a single

Table 1
Data sources.^a

Year 1	Honors Physics	Teacher A	1 SG	1 WC
Year 1	College Preparatory Physics	Teacher B	1 SG	1 WC
Year 1	Advanced Placement Physics	Teacher B	1 SG	1 WC
Year 2	Advanced Placement Physics	Teacher B	1 SG	1 WC

^a 1 SG and 1 WC indicate one class section taught in small group format and one in whole class format. From least to greatest difficulty were College Preparatory, Honors, and Advanced Placement classes.

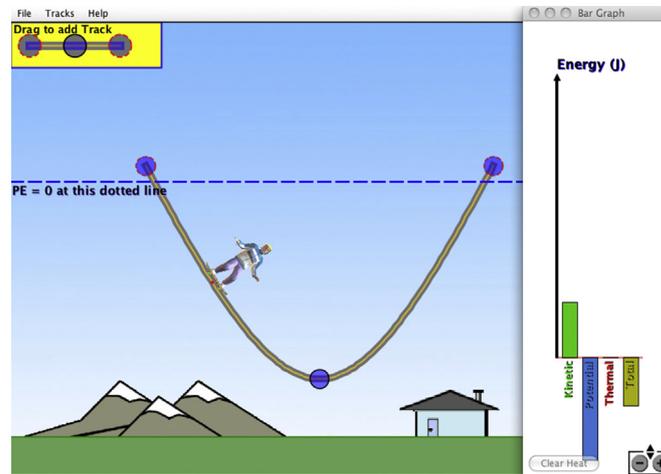


Fig. 1. The teachers chose PhET Energy Skate Park for the lesson, shown with the dotted GPE Reference Line and the Energy Bar Graph turned on. (<http://PhET.colorado.edu>).

computer to project the simulation onto a screen in front of the class and facilitated a whole class discussion as students worked through the activity sheets. In the small group condition, multiple computer stations were used with 2–4 students to a computer; they were allowed to engage in hands-on exploration and small group discussion guided by the same activity sheets as in the whole class condition while the teacher circulated among the groups. In both conditions, the teacher began by introducing the computer activity to the whole class, though the extent of this introduction varied. In both conditions, the teacher was available for questions the entire time the simulation was in use. Other than the constraints provided by the technological set-up, the activity sheets, the simulation condition (whole class or small group) and the data-collection needs of the study, teachers were free to conduct their classes as they saw fit and were encouraged to use the best teaching strategies they could devise for each situation. Control for time on task was implemented by using the same activity sheets and the same number of class periods to cover the material within each matched set. The lesson plans and activity sheets were developed by the teachers and reviewed by the authors. (Though early versions of the materials were inspired by sample lesson plans from the simulation website, <http://PhET.colorado.edu>, the final lesson plans and activity sheets were largely the construction of the teachers who participated in this study.) The pre/post tests were developed jointly by the teachers and research team and consisted of transfer questions that were not directly addressed during instruction; this was to minimize the possibility of the teachers' teaching to the test and also because the desire for these studies was to measure conceptual rather than rote learning. These tests were administered immediately before and after the instructional portion of the lesson sequence.

4.4. Gravitational potential energy lesson sequence

The teachers selected a simulation ahead of time from freely available online sources, Energy Skate Park at <http://PhET.colorado.edu> (Perkins et al., 2006). See Fig. 1. This is a sophisticated simulation developed through a series of formative evaluation trials. The track can be added to or reshaped, the skater placed anywhere in the scene and released, and the simulation run to see how the skater would respond under the influence of gravity (with or without friction). A sample page from an activity sheet and a sample pre/post test are provided in Appendices A and B.

After the whole class introduction, the lesson plan included five minutes of free exploration of the simulation (either in whole class or in the small groups) before students began work on their activity sheets. The activity sheets then supported students as they engaged in an exploration of the skater's motion, the changes in his/her potential, kinetic, thermal, and total energy with time, and the relationships between those changes (2nd order relationships). Depending on the level of the physics class, some of the activity sheets asked students to write their predictions for what would happen if the simulation were run for certain specified scenarios. All of the activity sheets asked whether the total energy of the skater could ever equal zero and requested a written explanation for the answer. All activity sheets explicitly included the instruction to turn on the Gravitational Potential Energy (GPE) Reference Line (see Fig. 1) and to move it around. Also included was an instruction to turn on the animated Energy Bar Graph, which showed clearly when the potential energy of the skater took on negative values. Late in the lesson were an exploration of a pre-set track configuration that included a full loop and an exploration of an "outer space" setting of the simulation that set gravity to zero. The classes ended with the post test.

5. Exploratory quantitative results

An exploratory study (Stephens, 2012) of learning gains via identical pre/post tests served to raise questions that we will attempt to address in the qualitative analyses that constitute the bulk of the paper. Scores were tabulated from multiple-choice and short answer questions that addressed conceptual issues. ANOVAs³ were used to compare the pre/post gains of the whole class and small group students within each matched set.

³ ANCOVAs completed recently yielded similar results.

Table 2
Gravitational PE multiple choice/short answer transfer question pre/post gains (Stephens, 2012).

	WC Gains			SG Gains			t-Value	Sig.	Cohen's d
	N	Mean	SD	N	Mean	SD			
CP	11	0.26	0.20	14	0.25	0.24	0.097	0.924	0.04
HP	20	0.22	0.21	19	0.09	0.15	2.221	0.033 ^a	0.71
AP	23	0.10	0.12	21	0.02	0.11	2.368	0.023^b	0.71
AP	21	0.09	0.16	21	0.07	0.10	0.506	0.616	0.16

Boldface indicates the larger mean gain within each matched set. In order of increasing difficulty were CP, HP, then AP course levels.

^a Significant difference in gains in favor of the whole class condition. This row is shaded to indicate the possibility that unanticipated events may have had a disproportionate effect on the small group condition; those students were encouraged, but not required, to finish their activity sheets in a single period while students in the whole class condition were not given this explicit suggestion. Both classes actually spent two days on the activity sheets.

^b Significant difference in gains in favor of the whole class condition.

Seven of eight class sections that met the criteria for matched-set analysis for the Gravitational Potential Energy sequence had statistically significant pre/post gains on multiple choice and short answer transfer questions, $p < 0.025$, effect sizes ranging from $d = 0.45$ (small) to $d = 1.40$ (large).

The results of ANOVAs comparing the pre/post gains of the two conditions are summarized in Table 2. In order of increasing difficulty level, there were College Preparatory Physics (CP), Honors Physics (HP), and Advanced Placement Physics (AP) classes. Gains are expressed as percentages of a perfect score.

As can be seen in Table 2, the effect sizes for the differences between whole class and small group performance on transfer questions were negligible in two of the comparisons, while in two others, the differences in gains reached significance in favor of the WC condition with a medium effect size. In one of these comparisons, there was a possible confounding factor with a slight variation in instructions that the teacher gave the two classes, but in the second comparison there was no obvious reason for the lower scores in the SG condition other than difference in condition. Because of this and the presence of significant effects in only some of the classes, we will be conservative in our description of the overall result: no advantage detected for students in the SG condition. But this result was still surprising given that both teachers had said at the time that the small group students appeared to be learning more.

These results from use of a sophisticated simulation agree with exploratory results from a related study on a Projectile Motion lesson sequence in which students used a very simple simulation and animations (Stephens, 2012). In that study, also, small group students showed no pre/post advantage over students who had experienced the materials solely in the context of whole class discussion. (See Fig. 2). The results are also consistent with two smaller studies of which we have since become aware, each of which showed no significant difference within a single pair of classes with lesson formats similar to the present study (Smetana & Bell, 2009; Wu & Huang, 2007).

5.1. Discussion of exploratory results

These results raised questions for us and for the teachers. At a follow up meeting with the teachers after the second year when they were shown the pre/post results, they expressed astonishment that the small group students had not outperformed the whole class students.

Given the nature and sizes of the samples, we did not attempt to conduct a comparison across all eight Gravitational PE classes but only to compare each small group Gravitational PE class with its matched whole class discussion class; in other words, to conduct four comparative analyses. The pre/post results have motivated us to conduct in-depth qualitative studies of these classes using videotape and activity sheet analyses. There are many lenses that could be applied to such analyses. Our general aim in undertaking the studies is to investigate student interactions with simulations and teacher support of those interactions; this constitutes our main focus. In particular, a review of observation notes and discussions with teachers suggested that there might have been interesting differences in student and teacher responses to, and interactions with, certain visual features of the simulation.

6. Qualitative analysis and results

6.1. Introduction to the qualitative study

The overall motivating question for these studies is whether there might be different strengths and weaknesses in the whole class and small group formats regarding student and teacher interactions with simulations. In view of the apparent lack of advantage with respect to pre/post gains for students who had used the gravitational PE simulation hands-on as compared with students in the matched whole class discussions, we also wish to shed light on this question: Why did the whole class format produce gains as strong as those of the small group format in these classes? Our hope was that videotape and activity sheet analysis would enable us to generate viable hypotheses to address these related questions. Because we wanted to approach this analysis as free as possible from preconceived ideas about what was happening in the classroom discussions, we employed elements of grounded theory to develop fresh analytical methods, as described below.

We used two types of data sources to investigate each matched set. The results of videotape analysis of the classroom discussions allow us to develop a picture of what an individual hypothetical student could have been exposed to in each class. In this analysis the video camera can be viewed as a proxy for an individual student; that is, the camera took the viewpoint of a hypothetical student in that classroom and recorded what an individual student might have seen and heard. In whole class discussion, the camera took a fixed position in the classroom and so captured some of the limitations likely to be experienced by student participants in that mode; at any position, some comments from fellow students were likely to be inaudible and accompanying hand gestures could be difficult or impossible to see. In classes in which the small group format was used, at the point at which the students moved into small groups, the camera moved to one of the groups also. Although fewer students were visible on camera than in the whole class condition, the videotape again recorded what an individual student in that group could have seen and heard, including occasional interactions with students from other small groups and with the teacher. In

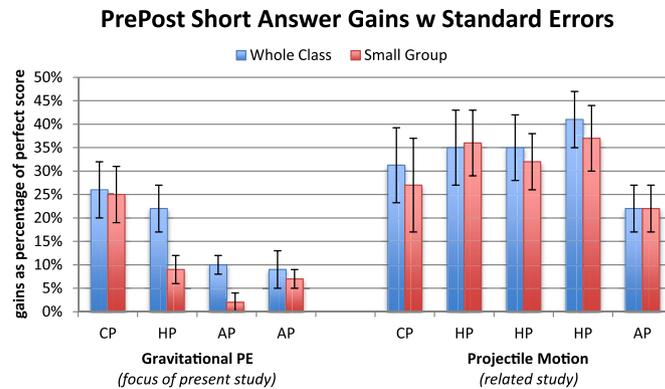


Fig. 2. Multiple choice/short answer pre/post gains (Stephens, 2012). Exploratory pre/post results from the Gravitational PE lesson sequence indicated a slight trend in favor of the whole class condition. This was consistent with exploratory results from a related study of a different lesson sequence; in no comparison was there a significant advantage for the small group students. AP courses were most advanced, CP courses least.

both conditions, we used videotape analysis to identify attempts by students and teacher to support the recognition and use of visual features within the simulation, and to support the understanding of conceptual issues.

Student activity sheets provide a different lens onto the students' experiences. This data source includes work from almost all of the students in the classes and is not restricted to students who spoke on camera. However, student drawing and writing abilities varied widely and some activity sheets were difficult to interpret or were not completed. Results of activity sheet analysis will be used to provide an estimate of how many students actually used certain visual features in their own thinking, as evidenced by their written and drawn answers to the relevant activity sheet questions.

6.1.1. Videotape analysis

We began by using a constant comparative method to identify key behaviors observable in videotapes and transcripts of four of the classroom discussions that occurred during use of the PhET Energy Skate Park simulation. Observation categories developed from this procedure were honed in an iterative process, along with coding criteria for assigning video segments to these categories. This honing process constituted a major part of the effort involved in this study; criteria were developed, applied to fresh transcript sections, then refined until the observation categories and their coding criteria stabilized. Finally, the criteria for the stable categories were used to code extended transcript selections from all eight transcripts, as described below. The coding results for each transcript can be thought of as providing one estimate for what an individual student could have been exposed to during the course of that lesson.

6.1.1.1. Example of videotape code map. Transana Transcription software (Woods & Fassnacht, 2007) was used when coding the videotapes. Short annotated excerpts from two Transana Code Maps are in Fig. 3. Each transcript is represented with time running from left to right with minute markers along the top. All portions of the transcript that addressed the topic of the lesson were segmented into utterances, usually one to three sentences that expressed a single idea or contribution to the discussion (as in McNeill & Pimentel, 2009). An individual's speech could be segmented into one or more utterances depending on how many ideas were being expressed. Alternatively, a single segment could include one or more speakers if they were speaking in unison or overlapping each other to present a single idea. (Portions of the transcript that involved technical difficulties or classroom activity not directly related to the lesson were not segmented into individual utterances, as in minutes 24–27 in the first Code Map excerpt below.) Each segment was examined to identify evidence that would assign it to one or more observation categories (described in Sections 6.2 and 6.3). If such evidence was identified, that segment was assigned a code for each suitable category, indicated in the code map by color blocks below the portion of the time line corresponding to that segment. The first Code Map excerpt in Fig. 3 is from a whole class discussion and the second is from a small group discussion.

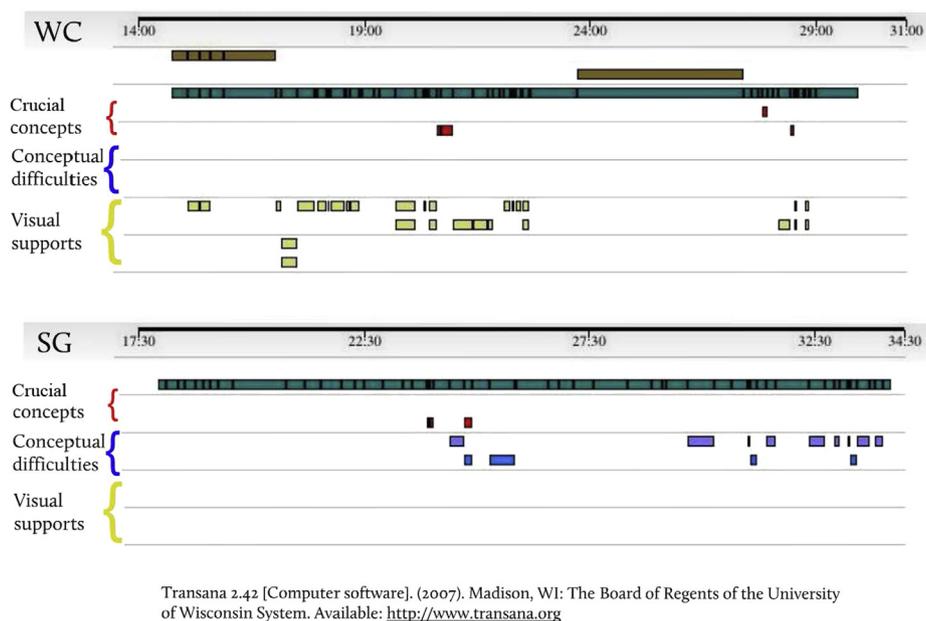
The small red blocks labeled Target Concepts in the code maps relate to Research Question 1; they indicate video segments in which discussion was occurring about certain concepts that had been identified as crucial for these students. The turquoise blocks labeled Conceptual Difficulties give a different perspective on Question 1; they indicate when conceptual difficulties were being addressed. Yellow blocks labeled Visual Supports relate to Research Question 2; they indicate that key visual features were being discussed or manipulated. An example of complete code maps from a matched set of discussions along with frequency data for each code is included in Appendix C.

Although the same amount of time was allowed for the lesson in each matched set, small groups varied in how much of that time they actually used for experimenting with the simulation and discussing it. However, this may have been because small groups could go at their own pace; some students might not have needed as much time in order to grasp the material. For that reason, the results of videotape coding will be expressed both as totals per discussion and as percentage of discussion time or average occurrence per unit time.

6.1.2. Activity sheet analysis

An important additional issue in addressing Research Question 2 is whether students actually recognized and used the key features in their thinking, although this question is more difficult to address than whether the features were discussed or manipulated. Where videotape analysis allows us to investigate aspects of what was occurring in the public space of classroom discourse, we also want to know what was appropriated and utilized by individual students in their own thinking. The pre/post tests consisted of transfer questions and were not very amenable to the kind of analysis that was needed. However, the activity sheets had been designed to help support student recognition and use of these features, and also to help detect whether this had occurred. In addition, the activity sheets gave an opportunity to look at the work of almost every student in a way that videotape analysis could not.

Videotape Code Maps



Transana 2.42 [Computer software]. (2007). Madison, WI: The Board of Regents of the University of Wisconsin System. Available: <http://www.transana.org>

Fig. 3. Excerpts from video code maps that illustrate results of qualitative analyses of a whole class (WC) and a small group (SG) discussion. The long teal bar near the top of each map indicates when the simulation was running.

We began our analysis by using a constant comparative method to code student responses to selected questions in a stratified sample of 30 activity sheets. Questions were selected that 1) addressed the crucial concepts that the key visual features were thought to support; 2) asked for open-ended written and drawn answers; 3) were not appreciably different among the activity sheets for the three physics levels; and 4) appeared to have elicited drawing and writing from most of the students in the sample of 30 sheets. Coding categories developed from this procedure were honed in an iterative process and coding criteria for those categories developed and refined. The refined criteria were used to code student responses to the selected questions on all 135 activity sheets for which there were legible answers (out of 150 total sheets). All coding was done blind to whole class or small group condition.

Examples of coded answers are provided in [Appendix A](#).

6.2. Question 1: To what extent were crucial physics concepts dealt with by teachers and students in the whole class and small group conditions?

We address this question in two ways, through coding the discussion for a) mention of specific concepts we had identified as crucial and b) evidence of any student conceptual difficulties (whether or not they were about concepts we had identified as crucial) and the responses made to those difficulties. Our intention was to obtain two different estimates of the amount of engagement these students had with the physics concepts that were crucial for them. For each of these sub-questions, we discuss the coding criteria used to identify evidence, followed by a summary of the results of videotape analysis using those codes.

6.2.1. Question 1a: To what extent do students and teachers engage in discussion about certain crucial concepts while working with the simulation?

From classroom observations of pilot lessons using this simulation the previous year, we had identified certain concepts that appeared not only to have been problematic for many of the students, but to have been crucial in allowing them to interpret the behavior of the simulation and its graphs and charts. Although the teachers had identified the idea of an arbitrary zero point (or height) for gravitational potential energy as having been a particular stumbling block, our observations led us to believe that an even more fundamental concept was posing difficulty for these students—the concept of the very existence of negative energy quantities. Although the simulation has valuable affordances for working with this concept (using visual features that will be explored in Research Question 2), it appeared to us that the lack of a concept of negative energy had constituted a block for the students in being able to interpret or work with these features in the simulation. Perhaps related to this, we had observed students questioning the existence of a total energy of zero for any system in the real world. The amount to which negative or zero energy quantities had been discussed in pilot classes had varied; in some classes the existence of such quantities had not been discussed at all. One forte of small group work has been presumed to be the fact that students have greater opportunity to raise questions. We wondered whether there would be a difference in the amount of attention paid to these two concepts during small group work as compared to whole class discussions in the lesson sequences to be observed for this study.

6.2.1.1. Coding criteria for Question 1a. To evaluate the extent to which these two crucial concepts were addressed during small group and whole class discussions, the following codes were used in videotape analysis.

Code: Student or teacher mentions possibility of total energy of some system being zero.

Code: Student or teacher mentions possibility of some kind of energy value being negative.

Total time and percentage of time spent on such discussion was noted for each videotaped discussion.

6.2.1.2. *Coding results for Question 1a.* The results in Table 3 can be used to estimate what an individual student in the position of the camera could have been exposed to in each class during the discussion that accompanied use of the gravitational potential energy simulation and activity sheet.

The percentage of discussion time coded for these concepts is shown in Table 3. Notably:

- Discussion coded for these two concepts ranged from 2% to 4% of discussion time in the small groups on camera and from 2% to 10% in the whole class discussions;
- The four small groups spent less time on the lesson, not because less time was allowed but because they chose to finish early, thinking they were done with the activity. Therefore, the *total* amount of time coded for these crucial concepts was substantially less in these small group discussions than in the matched whole class discussions, ranging from less than half a minute to a little over a minute in the small group discussions and from a minute to over four minutes in the whole class discussions.

In general, there was little discussion devoted to what a total energy value of zero or a negative energy value might mean, not exceeding 4½ minutes in any of the observed discussions. The smallness of these numbers was surprising, given that the animation provided important potential affordances for developing the concepts, including the two features shown in Fig. 1, and given the fact that these students, reminiscent of those in the pilot study, occasionally expressed frustration concerning these ideas. (This will be described further under Question 1b.)

Note that these concepts were not the only concepts discussed in these classes and are not the only important concepts necessary for students to understand the material. However, as described above, these had appeared to constitute a block to acquiring the other concepts of the lessons. The evidence described here does not suggest an advantage for the students in the small group condition regarding a chance to address these stumbling blocks. Even if the quality of discussion had been much higher in the small groups than in the whole class discussions, it is doubtful that less than half a minute of discussion, as in the small group in the lower level class, would have been sufficient to explore the concepts of zero or negative energy. If there was any trend within this sample, it was in favor of the whole class discussions, although no statistical analysis was attempted.

6.2.2. *Question 1b: To what extent did teachers and students attempt to respond to conceptual difficulties and misconceptions during work with the simulation?*

During pilot lessons, we had observed students expressing frustration and confusion concerning the two concepts described above and other conceptual matters. This had occurred in both whole class and small group discussions. In addition, we had observed students voicing misconceptions when they were not aware of experiencing any difficulty. There had been considerable variation in the extent to which classroom discussion, either whole class or small group, had addressed misconceptions or responded to expressions of conceptual difficulty. A presumed forte of small group work has been that individual students would feel more comfortable expressing their difficulties; however, students have varying ability and willingness to recognize and respond to the difficulties of their peers. We wondered to what extent student difficulties would be responded to in the small group and whole class discussions.

6.2.2.1. *Coding criteria for Question 1b.* Videotape segments that fit either or both of the following codes were coded as attempts to respond to a conceptual difficulty or misconception.

Code: *Response to conceptual difficulty:* Classroom activity following a student expression of conceptual difficulty was considered a response if it bore some relationship to the expressed difficulty.

Code: *Response to misconception:* Classroom activity was considered a response to a misconception if it appeared to be an attempt by teacher or student to address a misconception (sometimes pre-emptively; the misconception itself need not be in evidence).

We estimated the amount of discussion time spent on these responses by totaling the amount of videotape time assigned each code. No attempt was made to separate these into teacher and student responses; many responses were in the nature of joint discussion with overlapping comments.

6.2.2.2. *Coding results for Question 1b.* The results in Table 4 can be used to estimate what an individual student in the position of the camera could have been exposed to in each class during the discussion that supported use of the gravitational potential energy simulation and activity sheet.

Table 3

Discussion about 2 crucial concepts/length of discussion = percentage of discussion time.^a

Class	Teacher	Whole class format	Small group format
Yr 1 CP	Teacher B	4.32 min/42.42 min = 0.10	0.40 min/23.90 min = 0.02
Yr 1 HP	Teacher A	2.85 min/62.03 min = 0.05	0.75 min/29.23 min = 0.03
Yr 1 AP	Teacher B	0.92 min/41.10 min = 0.02	0.99 min/32.32 min = 0.03
Yr 2 AP	Teacher B	2.58 min/41.71 min = 0.06	1.16 min/28.95 min = 0.04

^a Results expressed in minutes, *not* in minutes and seconds. Boldface indicates the larger percentage in each matched set. CP = College Prep; HP = Honors Physics; AP = Advanced Placement.

The percentage of discussion time coded as attempts to respond to misconceptions and other conceptual difficulties is shown in Table 4. Notably, according to our codes:

- The whole class discussions spent a *greater percentage* of time in these responses than the matched small group discussions in 3 out of 4 comparisons.
 - From 4% to 11% of the time in small group discussions was coded as addressing such difficulties while from 4% to 23% of the time in whole class discussions was coded for this.
- The *total* amount of time spent on such discussion was longer in the whole class discussions in 3 out of 4 comparisons, coded portions ranging from 1 to 4 min in the small group discussions and from 2 to 14 min in the whole class discussions.
 - In the lower level physics classes the difference was considerable. In the medium level classes (HP), coded portions in the whole class discussion were 4× as long as in the matched small group discussion, while in the lowest level CP classes coded portions in the whole class were over 6× as long.

While one of the highest-level (AP) small groups spent longer addressing conceptual difficulties than did the matched whole class discussion, this was not true of the other AP comparison. Overall, there did not appear to be an advantage for those in small groups in terms of the amount of discussion time spent on conceptual difficulties and this was particularly true of the mid and lower level class discussions examined here.

6.2.3. Summary of results for Question 1

In three out of four matched sets, we coded more time spent on crucial concepts and more time spent on addressing student conceptual difficulties in the whole class discussions. This is consistent with the results of the earlier pre/post analysis and suggests a contributing factor as to why the whole class students had gains on conceptual transfer questions as high as those of the small group students.

6.3. Question 2: To what extent were key visual features of the simulation recognized and used in the two conditions?

Certain visual features in the simulation offered direct support for understanding the concepts we had identified as crucial. For instance, use of the Gravitational Potential Energy (GPE) Reference Line in conjunction with the animated Energy Bar Graph could produce a clear visual indication of the presence of negative energy quantities as the bars on the bar graph dropped below the x -axis. We had observed students noticing and commenting on this during pilot lessons. Likewise, for any set-up that resulted in a Total Energy (TE) of zero for the skater (Gravitational Potential Energy + Kinetic Energy + Thermal Energy), the bar representing TE in the bar graph disappeared altogether and remained absent no matter what the skater did on the track. However, during our classroom observations of the pilot lessons, we had noticed that teachers varied widely in how much they focused on these features. At times, we observed students experiencing difficulties that appeared to be more *perceptual* than conceptual, where they misinterpreted the meaning of a visual feature or failed to find it at all. On the other hand, we observed small group students in some groups helping each other identify and use these interactive features. We wanted to know how many of these episodes of teacher or student support for perceptual features occurred in the two conditions and we also wanted to gain an estimate of how much students were actually able to use the features in their thinking.

6.3.1. Question 2a: To what extent did teachers and students support the recognition, use, and interpretation of key visual features of the simulation?

We first used video analysis to compare how many episodes of teacher or student support for the key visual features could be identified in small vs. whole class discussions.

6.3.1.1. *Coding criteria for Question 2a.* Visual features in the PhET Energy Skate Park simulation that appeared to have been key for students during pilot lessons were the Gravitational Potential Energy (GPE) Reference Line and the animated Energy Bar Graph (see Fig. 1).

Code: Student or teacher supports use and/or interpretation of a key visual feature or relationship in the simulation.

Here, by “interpretation of a feature,” we mean the interpretation of its meaning, the development of some degree of understanding, as opposed to attaining rote knowledge of the feature or the ability to recreate a visual aspect through mimicry.

This code was assigned to a video segment when a student or teacher was observed making one or more of the following moves:

- 1) **Selectively pointing out** some aspect of the key visual feature or relationship as part of an apparent attempt to help students use it or interpret its meaning;
- 2) **Giving a hint** to encourage use or interpretation of the meaning of the key visual feature or relationship;

Table 4

Response to conceptual difficulties/length of discussion = percentage of discussion time.^a

Class	Teacher	Whole class format	Small group format
Yr 1 CP	Teacher B	6.15 min/42.42 min = 0.14	0.89 min/23.90 min = 0.04
Yr 1 HP	Teacher A	14.05 min/62.03 min = 0.23	3.35 min/29.23 min = 0.11
Yr 1 AP	Teacher B	3.72 min/41.10 min = 0.09	1.58 min/32.32 min = 0.05
Yr 2 AP	Teacher B	1.79 min/41.71 min = 0.04	3.12 min/28.95 min = 0.11

^a Results expressed in minutes, *not* in minutes and seconds. Boldface indicates the larger percentage in each matched set. CP = College Prep; HP = Honors Physics; AP = Advanced Placement.

- 3) **Gesturing in the air or over the display** to indicate the key visual feature or relationship as part of an apparent attempt to help students use it or interpret its meaning;
- 4) **Asking a question to prompt use or interpretation** of the meaning of the key visual feature or relationship;
- 5) **Suggesting a manipulation of the simulation** to assist with use or interpretation of the meaning of the key visual feature or relationship;
- 6) **Pointing out a limitation** to interpreting the meaning of the key visual feature or relationship.

Generally when any one of the six moves was undertaken in an attempt to provide visual support, it was considered a single visual support 'episode,' which was our unit of measurement. If the teacher or student simultaneously engaged in more than one of these moves, such as *selectively pointing out a key visual feature while simultaneously asking a question to prompt students to interpret its meaning*, this was counted as a single support episode. In a long series of support moves, a pause for response or a shift in tactics (asking a different prompting question, for example) was considered to demarcate between episodes. However, if the same move was repeated several times in a row, it was counted as a single episode.

6.3.1.2. *Coding results for Question 2a.* The results in Table 5 can be used to estimate what an individual student in the position of the camera could have been exposed to in each class during the discussion that supported use of the gravitational potential energy simulation and activity sheet.

The intention was to identify the amount of support used to address student *perceptual* and other difficulties in making effective use of the key visual features described above, which were intended affordances of the simulation for this lesson sequence. Either a teacher or student could employ these support moves.

The frequencies of visual support episodes are given in Table 5.

- Rates of visual support episodes ranged from 10 to 21 per hour for the small group discussions and from 25 to 52 per hour for the whole class discussions.
- Total numbers of episodes ranged from 4 to 10 per small group discussion and from 17 to 37 per whole class discussion.
- In no comparison did the small group discussion show an advantage in visual support for the key features, either in rate or in total number of support episodes.

Note that episodes of mutual support *were* counted; it was not required that the person engaging in support be acting as an expert, only that the move appeared intended to help other students in addition to the supporter. This is because we believe that students can help each other even when they are all at a similar level of expertise and we wished to identify such support whenever it occurred. We expected to see this happen more often in small groups than it did. Interestingly, many of the student-on-student small group episodes occurred when the teacher stopped by.

6.3.2. *Question 2b: Did students recognize and use key visual features of the simulation?*

Although the whole class discussions appeared to give rise to more visual support episodes than small group discussions, this does not necessarily mean that students benefited more in the whole class situation. They might have been able to experiment with and learn from the key features more easily in the hands-on situation of the small groups, rendering support episodes less necessary during the small group work. We wondered whether written and drawn work of the students would give evidence for their recognition and use of the key features and, if so, whether there would be a difference in amount of such evidence according to whole class or small group condition. The activity sheets completed by all students during their work with the simulations proved amenable to this analysis. A constant comparative method was employed to identify and hone coding categories, and all activity sheets were then re-coded with the codes below. All coding was done blind to whole class or small group condition.

6.3.2.1. *Coding criteria for Question 2b.* The visual features identified as key were the movable Gravitational Potential Energy Reference Line and the animated Energy Bar Graph (Fig. 1). Two crucial concepts, the possibility that energy could take on negative values and the possibility that the total energy of a system could equal zero, could both be explored by coordinated use of these two features, as described in Section 6.3 above. The activity sheet questions that were determined appropriate for analysis (as described in Section 6.1.2) directly asked about these two concepts. As students tried to describe their understandings of the concepts, they frequently mentioned the key features or indicated the features in their drawings.

Student written and drawn answers were coded separately for the following, which were considered to be plausible indicators of 1) use of the movable Reference Line in their thinking, 2) use of the Energy Bar Graph, and 3) coordinated use of the two of them together. (Examples of student answers and the corresponding codes are given in Appendix A.)

Table 5
Episodes of support for key visual features/length of discussion = episodes per hour.^a

Class	Teacher	Whole class format	Small group format
Yr 1 CP	Teacher B	37/42.42 min = 52 /hour	4/23.90 min = 10/hour
Yr 1 HP	Teacher A	26/62.03 min = 25 /hour	8/29.23 min = 16/hour
Yr 1 AP	Teacher B	17/41.10 min = 25 /hour	10/32.32 min = 19/hour ^b
Yr 2 AP	Teacher B	19/41.71 min = 27 /hour	10/28.95 min = 21/hour

^a Boldface indicates the larger rate within each matched set. In order of increasing difficulty level: CP = College Prep; HP = Honors Physics; AP = Advanced Placement.

^b The students in the small group engaged in support for other features that they adapted in place of key features they could not find. This number represents the amount of support for key features only, but under-represents the amount of visual support these small group students actually provided each other.

Code: Student refers to the GPE Reference Line in a way that implies that the line is movable. (See Appendix A.)

Code: Student answers contain evidence for use of at least one of 3 concepts supported by the animated Energy Bar Graph:

- Amount of gravitational potential energy (PE) and kinetic energy (KE) change in opposition to each other; as one increases, the other decreases;
- The total energy (TE) reading for the skater can change (even when the physical configuration has not changed; in other words, when the reference line has been moved);
- PE and/or TE quantities can become negative.

Code: Student answer contains evidence for use of the key relationship supported by coordinated use of the two key features:

- TE and/or PE depend on position of the reference line, where the implication is that both position of the line and the energy amount can change.

Each student's written work was assigned either a 1 or a 0 for each of these codes. (See Appendix A.)

6.3.2.2. *Coding results for Question 2b.* Table 6 and Fig. 4 show the percentage of students in each section whose written and/or drawn work exhibited evidence for use of either of the two key visual features and/or the visual relationship between them, as described above. In Table 6, note that the comparison of results is laid out with small group student data listed below the matched whole class data rather than to the right as in Tables 2–5. Because the same data were scored along all three dimensions, the results are not added across dimensions.

In Fig. 4, each group of 3 bars represents a single class analyzed along 3 binary dimensions (labeled 1, 2, 3 in Table 6), where each bar represents the percentage of students in that class exhibiting evidence for use of a single feature or relationship. For instance, the first bar in each group represents the percentage of students who referred to the reference line in a manner consistent with its being movable.

Notably:

- In every instance in which the teacher facilitated whole class discussion about the activity sheet questions (did not inadvertently skip them; see Table 6), a greater percentage of students in the whole class condition exhibited, in their written and drawn work:
 - Evidence for using the GPE Reference Line in their reasoning,
 - Evidence for using one or more concepts supported by the Energy Bar Graph,
 - Evidence for using the relationship between position of the GPE Reference Line and amounts of PE and TE.
- The only small group students who showed evidence on their activity sheets for having used the key features were Advanced Placement students;
 - No student in any of the Honors Physics or College Preparatory small group classes showed evidence along any of the three dimensions analyzed—they showed no evidence in their written or drawn work for having used either of the key features or the relationship between them.

Comparisons of the small group and whole class results for all students who answered the questions on the activity sheets suggest no advantage for the small group students over the whole class students in being able to incorporate use of key visual features and relationships in their thinking, at least as evidenced by their written and drawn responses. This is consistent with the results of activity sheet analysis in the related study (Stephens, 2012) for a somewhat different kind of activity sheet.

6.3.3. Summary of results for Question 2

In videotape analysis, in no whole class/small group comparison did the small group discussion analyzed show an advantage in visual support for the key features, either in rate or in total number of support episodes. Activity sheet analysis yielded even stronger results: we were able to identify no written or drawn evidence for use of the key features by any of the students in the intermediate or lower-level small group classes. Taken together, these results suggest a second factor that could have contributed to the pre/post results: the frequency of visual support episodes occurring during small group discussions may not have been sufficient to allow these mid- and lower-level students to make use of the simulation's visual features in their own thinking.

Table 6

Performance on relevant activity sheet questions: percentage of students who exhibited the evidence described.^a

Class	Teacher	N	Lesson format	1) Evidence for use of GPE ref line	2) Evidence for use of concepts supported by bar graph	3) Evidence for use of key relationship
Yr 1 HP	Teacher A	20	WC	0.10	0.05	0.05
Yr 1 HP	Teacher A	18	SG	0.00	0.00	0.00
Yr 1 CP	Teacher B	11	WC	0.36	0.27	0.18
Yr 1 CP	Teacher B	13	SG	0.00	0.00	0.00
Yr 1 AP	Teacher B	13	WC	0.15	0.23	0.08 ^b
Yr 1 AP	Teacher B	18	SG	0.33	0.44	0.22
Yr 2 AP	Teacher B	21	WC	0.95	1.00	0.95
Yr 2 AP	Teacher B	21	SG	0.81	0.95	0.48

^a Boldface indicates the larger percentage within each matched set.

^b Teacher inadvertently skipped the relevant portion of the activity sheet during whole class discussion.

Selected Activity Sheet Questions scored along 3 dimensions

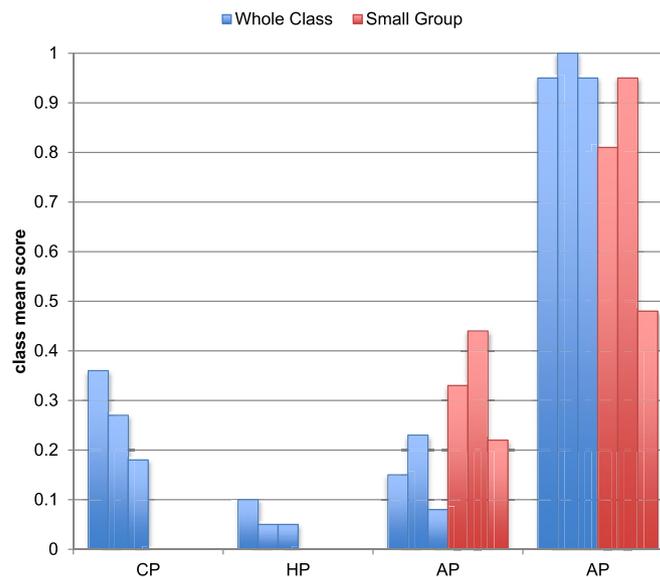


Fig. 4. Activity sheet work: Percentage of students exhibiting three types of evidence for use of key visual features. Coding was blind to condition. Each blue-red cluster represents a matched set of classes. Red bars for lower and medium level (CP and HP) small group classes are not seen—they are zero for all three types of evidence.

7. Discussion

Although almost all of the classes had shown significant gains on the pre/post short answer questions, the teachers were surprised that there appeared to have been no pre/post advantage for students in the small group condition. This was true even though small group participants had had the advantage of hands-on experience with the simulations, the time for each and every student to raise questions with group mates and with the teacher, presumably a more relaxed atmosphere in which shyer students could speak up, and the increased engagement the teachers reported for the small group work. Why did the hands-on keyboard small groups not do better than the whole class students? What strengths and weaknesses of the two lesson formats were suggested by videotape and activity sheet analyses? While there were many lenses through which we could have addressed these questions (argumentation, socio-cultural, etc.), our preliminary observations led us to focus on conceptual and perceptual issues directly related to the interactive visuals. The qualitative videotape and activity sheet analyses suggest factors that could help explain the pre/post results.

To what extent were crucial physics concepts dealt with by teachers and students in the whole class and small group conditions? Videotape analysis indicated that certain concepts that appeared important for these students were discussed very little in either condition, and gave no evidence for an advantage for the small group condition. First, analysis showed a greater percentage of discussion time spent on two crucial concepts in the whole class discussions in 3 of 4 matched sets of class sections. Second, our analysis showed a greater percentage of time spent on addressing student difficulties in the whole class discussions in 3 of 4 matched sets, and total time considerably longer, especially in the lower level classes, in which the whole class discussions were coded as spending 4× and 7× as long on student difficulties. These first two results together suggest that if there were advantages from having hands-on keyboards, these did not translate to more time spent on these conceptual issues. This could be one reason the small group students did not outperform the whole class students on the conceptual transfer questions on the pre/post tests.

To what extent were key visual features of the simulation recognized and used in the two conditions? First, all four of the whole class discussions analyzed here had more episodes where a teacher or student provided support for using the key visual features of the simulations than in the matched small group discussions that were analyzed. The whole class discussions also exhibited a greater frequency of such episodes. Second, activity sheet analysis revealed that in 3 of 4 comparisons, students in the whole class discussions exhibited more evidence for actually having used the visual features in their own thinking. Notably, the only small group students who showed any evidence for use of the features were in the Advanced Placement classes, the highest-level classes in the study. None of the students in the Honors or College Prep small group classes exhibited any evidence for use of the visual features along any of the three dimensions examined. These results suggest another hypothesis that could help explain the pre/post results: the frequency of visual support episodes occurring during small group discussions may not have been sufficient to allow these mid- and lower-level students to make use of the visual features in their own thinking.

The argument here is not that the small group work did not have benefits—it clearly did; small group students had pre/post gains almost, if not as, large as the whole class students. Rather, we argue that the whole class and small group formats appeared to have compensating strengths and weaknesses when it came to learning from the sophisticated physics simulation used here.

Our classroom observations and immersion in the transcripts, as well as additional case study analysis in Stephens (2012), suggest several grounded hypotheses for why the small groups did not do better than they did: there appeared to be a tendency in some groups to cut off conceptual discussion in the interests of time; some groups appeared to have a “get and report data” mindset that could have limited conceptual discussion; and visual support—even student-on-student support—appeared to cluster around teacher visits to the small groups. (An example that shows some clustering of analytical codes around teacher visits can be seen in Appendix D.) We hypothesize that

these characteristics were present enough in the small groups to have an impact on their pre/post performances, counterbalancing the advantages of the hands-on opportunities these students had with the simulation. But all of this raises the question of *why* the small groups exhibited these characteristics.

7.1. Hypotheses concerning group mechanisms

Going beyond systematically observed patterns and hypotheses grounded in the data, in this section we will form some additional hypotheses that, if true, could help explain the findings.

When small group students experienced a difficulty interfering with their progress on the activity sheet, they, unsurprisingly, tended to *call the teacher over*; thus, it makes sense that responses to conceptual difficulty would occur more frequently when the teacher was present. However, in the code maps, episodes of *student-on-student visual support* (that is, when students supported each other to recognize visual features in the simulation) also appeared to cluster around teacher visits (See [Appendix D](#) for one example). One possibility is that the teacher's presence may have helped focus the student discussion, including the student–student exchanges. Another is that there may have been socio-cultural factors at work; in small groups, we observed some students respond to the difficulties of fellow students with teasing or even belittling. However, when the teacher stopped by a small group, different dynamics may have been in play for the students; offering help to their fellows may have been more attractive or acceptable.

The teachers had anticipated that students would be more reluctant to raise their conceptual difficulties in whole class than in small group discussion. It was true that some whole class students were observed softly voicing their dissatisfaction or puzzlement to other students rather than raising their hands to ask questions. However, one of the teachers developed an interesting way to deal with this; she appeared to watch for such exchanges and then to repeat the murmured comment loudly and enthusiastically to the whole class. [Windschitl, Thompson, Braaten, Kang and Stroupe](#) refer to repeating a student comment as 'revoicing.' In these whole class discussions, this move appeared to validate the topic as worthy of discussion, and the students frequently responded with animated and engaged discourse. This support strategy may have helped offset any increased reluctance to speak up in the whole class setting.

There appeared to be a difference in the ways in which whole class and small group discussions responded to unexpected time pressure. It was our impression that the small groups had a tendency to respond to such pressure by "knuckling down" to the task at hand, trying to complete the activity sheets. One of the videotapes clearly shows students monitoring their time and cutting conceptual discussion short in order to maintain their progress through the sheet. During whole class discussions, on the other hand, the teachers clearly felt free, even impelled, to diverge from their lesson plans and expand the time on task when conceptual difficulties arose, even if this meant doubling the intended time (for both classes in the matched set) or abandoning their plans for equivalent time on task (which occasionally necessitated classes being dropped from our analyses). Another factor is that even though small group students were provided equivalent time in all the classes included in this study, they often chose to spend less time on the activity sheets than was spent in whole class, even in the absence of time pressure. This may have contributed to the quantitative pre/post results, but we believe we have also identified other factors that can contribute.

Studies have reported a variety of issues concerning the effective use of small group discussions in science classes, including that students can exhibit a low level of engagement with tasks ([Bennett et al., 2010](#)). In addition, there are several, more speculative explanations that might be able to account for part of our results. For instance, our impression was that some groups tended to view the activity in terms of finishing the worksheet in the shortest amount of time. On the other hand, we occasionally observed groups still actively trying to make sense when the bell rang. These could be described as ways that students *frame* the classroom activity, including *epistemological* frames ([Hammer, Elby, Scherr, & Redish, 2005; Redish, 2004](#)). Although our camera set-ups were not designed with this in mind, we would anticipate that an analysis of the small group discussions in terms of frames along the lines of Redish and Hammer ([Hutchison & Hammer, 2010; Redish, 2004, 2014; Rosenberg, Hammer, & Phelan, 2006; Scherr & Hammer, 2009](#)) might reveal a lot of moment-to-moment variability within, as well as an overall variability between, small groups. With a different camera set-up than the one we used, similar analyses could presumably be done for whole class discussions as did [Hutchison and Hammer \(2010\)](#). If the balance of, for instance, "completing the worksheet" versus "discussing ideas" epistemological frames ([Scherr & Hammer, 2009](#)) was appreciably different in the two conditions, this could suggest additional competing strengths and weaknesses that could have contributed to our results.

Many of the studies cited in the previous paragraph were of undergraduate science students and the focus was on epistemic frames, where most students had a goal of learning the material, and what varied was what "learning" and "understanding" meant to the different students. However, if a student's goal is to finish an activity in the shortest amount of time, it seems likely that the components of her frame for the activity are other than epistemic. [Pope \(2001\)](#) described students whose goals were to get good grades for reasons not related to actually acquiring knowledge. Reports from the high-school teachers in our study as well as follow-up interviews with students of these same teachers in a different semester ([Stephens, 2012](#)) indicated to us that students tended to enroll in these classes for reasons ranging from wanting to enter a STEM career, to wanting to be in a class with their friends, to wanting good grades as a step to a career for which they believed science was irrelevant. The effects of these differing student goals and consequent framing of the activities may have played out differently in whole class and small group contexts.

A competing hypothesis to consider is whether the activity sheets were poorly designed for use in small groups. However, the teachers spent considerable time on the design and actually designed the sheets with the small groups in mind, partly because they were more accustomed to using activity sheets for simulations in that way. Although the sheets used in the two contexts were identical, they seemed to function somewhat differently. For instance, at times teachers appeared to feel freer to allow departure from them in the whole class format than students did when in the small group format. Regardless of their design, their use may have been framed differently.

The final hypothesis we would like to suggest concerns use of scaffolding strategies in the two discussion formats. As mentioned earlier, authors such as [Clement & Ramirez \(2008\)](#), [Chang and Linn \(2013\)](#), and [Windschitl et al. \(2012\)](#), and others have argued for the importance of focused scaffolding strategies when teaching complex models in science, consonant with theories of model-based learning. Scaffolding strategies are designed not only to engage students in model-based reasoning but to support students in doing the reasoning. [van Zee and Minstrell \(1997\)](#), [Windschitl, et al. \(2012\)](#), and others have documented moves that experienced teachers use to draw out and extend student contributions by asking them to elaborate on their initial contributions. Among the moves we observed the two teachers making in this study were asking students to elaborate and setting up 'special cases' in the simulations designed to surprise and engage students and to

provoke reasoning. A plausible hypothesis is that these strategies may have been easier to deploy broadly in whole class discussion, helping to explain the lack of advantage of small groups with respect to pre/post gains.

Our impression is that when small groups worked well, they could work very well, but that this was true of a minority of the groups. In general, the whole class discussions appeared to be of a richer quality. Some students in both situations appeared to experience a lack of engagement; however, small group students who were not engaged appeared generally to remain disengaged for the whole class, while even the most disengaged whole class students appeared to become engaged from time to time, especially when the teacher did something unexpected with the simulation.

7.2. Implications

7.2.1. Design implications

The results of this study suggest design considerations for educational physics simulations and further considerations for the design process itself. Design principles developed for educational animations and simulations, whether from dual coding theory (Mayer & Moreno, 2002) or iteratively developed by designers (e.g., Adams, et al., 2008b) may provide guidance in design only to a first approximation; successful design probably requires iterative cycles of testing and refinement and at least some of this testing probably needs to be done in the noisy environment of the classroom. (This is currently the practice of the PhET group, Concord Consortium, and other successful design teams; our results serve to underscore the importance of this last, rather expensive step in the design process.)

Although many physics simulations undergo iterative cycles of testing and refinement, they might benefit from trials in situations beyond the one-on-one and small group trials often employed. Although not as commonly done, trialing a sophisticated simulation in a whole class condition may suggest support materials that are needed to make more productive use of the simulation. This kind of trial may also suggest additional interactive features to facilitate whole class use. For instance, general design requests by teachers we worked with in these studies (requested for more than one simulation) have included:

- To provide the ability to mask arbitrary parts of a display so that teachers can set up novel scenarios and ask students to predict what will happen next;
- To provide the ability to save multiple starting conditions so that teachers can create and test set-ups beforehand and easily switch set-ups as desired.

We believe these suggestions are consistent with the design principles recommended by the authors cited above, although those principles are not as often applied to a whole class context. Not only would these provisions help focus the students' attention onto crucial concepts and features, but they are designed to promote active engagement by requiring students mentally to run the simulations in order to make predictions.

7.2.2. Pedagogical implications

Adams et al. (2008a), believe that in order for simulations to be highly effective, the student's interaction must be directed by the student's own questioning. (This lesson sequence used one of their popular simulations.) Our results suggest that students in small groups in a classroom environment may not always direct their own questioning or may not be able to. A more important factor than whether a student has hands on keyboard may be whether the student has "minds-on" (Duckworth, Easley, Hawkins, & Henriques, 1990), and at times, we suggest, this may be accomplished as well or better by using a simulation in the context of a whole class discussion. In the most animated discussions we observed, we believe the teacher was allowing the whole class students to guide the activity by their questioning, although she also primed some questions by causing the simulation to behave in surprising ways (ways that most of the small groups did not discover).

van de Sande & Greeno (2012) hypothesize that a student's understanding can be framed with some elements or concepts in the foreground, with others backgrounded or absent, and that which elements are foregrounded can vary. When a teacher stands next to a screen displaying a simulation and leads a discussion, it may tend to foreground, or focus students on, the simulation and the topic. A question for future research is whether these and other teacher moves served to foreground the simulations over the worksheets in the whole class situation.

We also suggest that the teachers in our study used a variety of strategies to foreground subtle but important visual features within the simulations and it may be that the teachers had a stronger hand in this in the whole class situation. However, our classroom observations suggest that many teachers may need more guidance provided along with simulations to help them identify which features and relationships are likely to be overlooked by students. Teachers may also need suggestions for making these features explicit.

The fact that the students in these whole class discussions matched or exceeded the performance of their small group peers implies that there were teaching strategies for promoting at least some of the active thinking and exploration that has been considered to be the strength of small group work. Now that we are aware of more factors that can impede or support learning, it is intriguing to contemplate what would happen if teachers were taught some of the supporting strategies we have identified (visual support strategies identified during videotape analysis listed in Section 6.3.1.1; large list of whole class discussion strategies for scaffolding simulations on our website, <http://www.umass.edu/teachingstrategies/>), and if they were informed about the importance of training students to use certain learning strategies when using simulations in small groups. Neither types of training were done during the present study, but we feel this is an important area for further research.

7.3. Conclusions

Teachers we have spoken to believe, sometimes quite strongly, that simulations are more effective when used by small groups. However, pre/post results had suggested a slight trend in favor of the students in the whole class condition, raising the question of why the small groups did not do better. Analyses of matched whole class and small group discussions during use of a sophisticated interactive physics

simulation revealed that in the whole class discussions analyzed here, there was 1) more time spent on crucial concepts; 2) more time spent addressing student conceptual difficulties; and 3) more episodes providing support for using key visual features of the simulations. In addition, analyses of student worksheets revealed that for students in the whole class condition, there was 4) more evidence for student use of the key features in their thinking. In fact, no small group students in the medium or lower level physics sections showed any evidence along several parameters for having utilized the visual features in their thinking. These factors may counterbalance or complement the natural advantage of small group formats for spontaneous interactions with the computer, and we recommend using a mixture of the two lesson formats.

The slight trends in observed gains suggest that research on larger populations might yield more significant results regarding an overall advantage for a particular discussion format. However, we suggest that a more productive line of research would be to investigate what mixture of the two formats might be optimal and at what points each of the discussion formats is best used. Our results suggest that there may be certain instructional situations (e.g., when the intention is to provide consistent support for interpreting and using information from onscreen visual elements) where there is an advantage to spending at least part of the time with a simulation in a whole class discussion mode.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants DRL-1222709 and DRL-0723709 awarded to John J. Clement. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix A. Sample from activity sheet to accompany whole class and small group discussions

4. Find the actual values of the different energies by clicking on the **SHOW PATH** button and letting the rider lay down several rounds of purple dots before clicking **PAUSE**. Now click on a purple dot at the top and at the bottom to display data that will help you to determine where each type of energy is the most. (You may need to hide the PE reference line in order to click on dots underneath it.) Record the values, **rounding to the nearest 10 Joules**:

Most KE: _____ Most PE: _____ Total Energy: _____

- Does $KE = \frac{1}{2} m v^2$? (show your calculation)
- Does $PE = mgh$? (show your calculation)
- What does total energy mean?



Click **CHOOSE SKATER** to find its mass

RESET. Turn on the energy Pie Chart and Bar Graph.

5. Without changing anything else, use the **CHANGE SKATER** button to explore changing the skater's mass. How does changing the skater's mass affect each type of energy?

Potential Energy:

Kinetic Energy:

Total Energy:

- Could either kinetic energy or potential energy ever be *less than zero*? (remember, you can move that reference line around...)
 - $KE < 0$?
 - $PE < 0$?
- Could total energy ever be less than zero? Explain your reasoning.

Fig. A1. Sample from Advance Placement activity sheet. Examples of coded student answers to Questions 6 and 7 are also in this appendix.

Examples of coded student answers to relevant questions

The phrases below are from written portions of student answers to relevant activity sheet questions, including Questions 6 and 7 in Fig. A.1. Drawn portions of student answers were coded separately. Each answer was coded along 3 binary dimensions, as exhibiting evidence or not exhibiting evidence for use of the concept or relationship.

Dimension One: Does student refer to the GPE Reference Line in a way that implies that the line is movable?

Yes = 1 (sufficient evidence)	No = 0 (insufficient or no evidence)
"When the object is below the ref line"	"When h is always negative"
"In our setup with the line at the bottom, TE stays the same."	"When not moving at $h = 0$ "
"If you put the zero bar at the top where he stops momentarily"	"If he's at the bottom, not moving"

Dimension Two: Do student answers contain evidence for use of any of 3 concepts supported by the animated Energy Bar Graph?

1) PE and KE change in opposition to each other, when one goes up, the other goes down; 2) TE reading for a system can change (even when the physical configuration has not changed; in other words, when the reference line has been moved); 3) PE and/or TE quantities can become negative.

Yes = 1 (sufficient evidence)	No = 0 (insufficient or no evidence)
"If you have no KE, then you have PE, and vice versa." (1)	"Only if he were on the ground not moving"
"TE can be zero when he goes lower than the reference line." (2) ^a	"TE can never change because energy is conserved."
"If PE becomes negative and is larger than KE" (3)	" $-PE > KE$ "

^a Why does this answer imply a change in TE? The default for the skater was a positive TE. Some students believed he could never have zero TE because "TE can never change." Although this answer implies some relationship between TE and the reference line, it is not explicit enough to satisfy Dimension 3; it does not indicate that the position of the reference line can be changed.

Dimension Three: Does student answer contain evidence for use of the key relationship supported by coordinated use of the two key features?

Key Relationship: TE and/or PE depend on position of the reference line, where the implication is that both position of the line and the energy amount can change.

Yes = 1 (sufficient evidence)	No = 0 (insufficient or no evidence)
"TE can be negative if zero bar is placed higher than track, neg PE > pos KE"	"TE can be negative when h is negative and object is below the ref line"
"Depends on ref line, $PE + KE < 0$ "	"If skater never goes above the zero line"
"If zero point very high and KE low enough"	"If negative PE > KE, if trial is under the curve"

These dimensions were used to identify evidence for student use of each of two visual features and/or the relationship between them. As such, they are not orthogonal dimensions; many—though not all—phrases and drawings coded Yes for Dimension Three were also coded Yes for the other two dimensions.

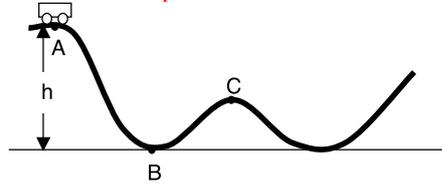
Appendix B. Gravitational Potential Energy Pre/Post Test (SAMPLE PAGE)

Teacher B Pre-Test Name _____ Date _____

Max short answer: 9

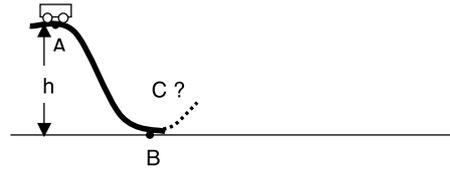
Max explanation: 2

1. Consider the roller coaster to the right. Assume no friction or air resistance. The cart starts from rest at point A and begins to roll to the right.



- a. Will the roller coaster pass over point C? Explain why or why not.
 N/A (Pretest ceiling effect)

- b. How high could point C be and still have the roller coaster pass over it? Sketch that here:



- c. Fill in the table with the cart's gravitational potential energy and kinetic energy, assuming the heights of the 3 positions are as indicated.

Each cell: 1/2 pt

	A = 10 meters	B = 0 meters	C = 5 meters
Gravitational Potential Energy	1000 J	0	500J
Kinetic Energy	0 J	1000J	500J

- d. How would these values change if there were friction present? How would the values change if the roller coaster were located on the Moon? Fill in the chart below indicating INCREASE (I), DECREASE (D) or SAME (S) as compared to part (c).

Each cell: 1/2 pt

		A	B	C
With Friction	Gravitational Potential Energy	same	same	same
	Kinetic Energy	(0)	decrease	decrease
Moon, no Friction	Gravitational Potential Energy	decrease	same	decrease
	Kinetic Energy	(0)	decrease	decrease

Fig. B1. Page from Advanced Placement pre/post test with correct answers in red.

Appendix C. Example of code maps from one matched pair of Gravitational Potential Energy discussions

Each time line in Fig. C1, running from left to right, represents 50 min of classroom videotape. Not all of that time was taken up by discussion. The solid teal bar (5th row down) indicates the portion of the timeline during which the discussion was focused on the simulations and associated activity sheets, while the bronze, navy, and gray blocks in the first four rows indicate the general nature of the classroom activity. A brief description of each code is on the left. (Fig. C1 is in color in the web version of this article.)

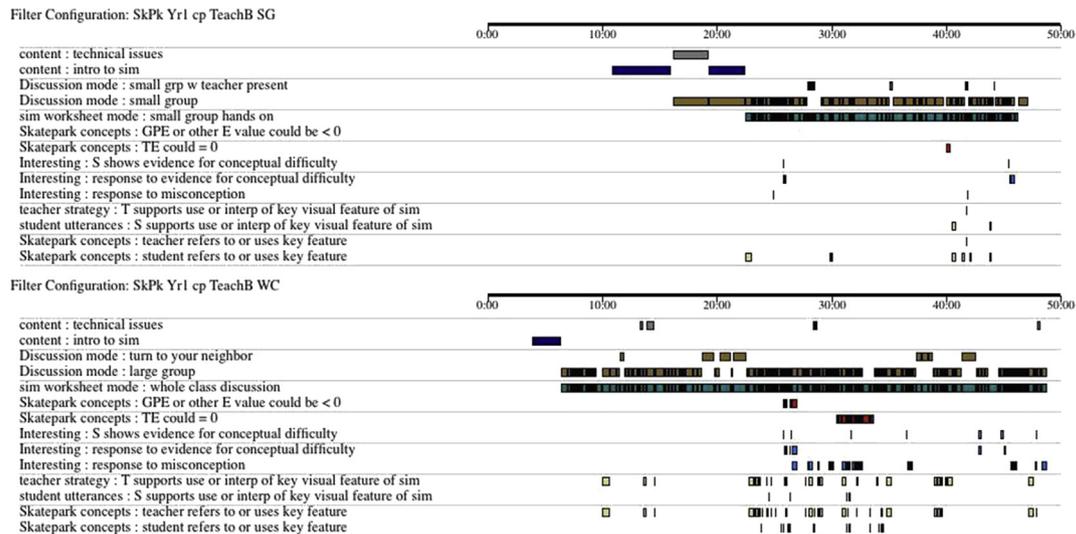


Fig. C1. Code maps for matched WC (top) and SG (bottom) discussions.

Table C1

Numerical representation of results represented in Figure C.1.

	Whole class format	Small group format
<i>Length of analyzed taped discussion</i>	42 min 25 s	23 min 54 s
<i>Research Question 1a: Discussion about two crucial concepts</i>	Total length: 4 min 19 s Percentage of discussion: 10%	Total length: 24 s Percentage of discussion: 2%
<i>Research Question 1b: Response to conceptual difficulties and misconceptions</i>	Episodes of difficulty: 8 Response length: 1 min 29 s Response to misc w no prior evidence of diff: 33 Length: 4 min 40 s Total: 6 min 9 s	Episodes of difficulty: 2 Response length: 41 s Response to misc w no prior evidence of diff: 2 Length: 13 s Total: 54 s
<i>Research Question 2a: Support for key visual features</i>	Percentage of discussion: 14% Total support episodes: 37 Teacher: 29 Student: 8 Avg: 52 per hour	Percentage of discussion: 4% Total support episodes: 4 Teacher: 2 Student: 2 Avg: 10 per hour

From Table C.1, it can be seen that, compared to the small group on camera, the matched whole class discussion had:

- Six times the *percentage* of discussion time (10.2% vs. 1.7%) spent on crucial concepts, (more than 10× the amount of *actual* discussion time);
- Several times the *percentage* of discussion time spent on addressing conceptual difficulties and misconceptions;
- Much greater frequency of support for using and interpreting key visual features.

The whole class discussion also lasted longer; this was true in many of the comparisons. In each classroom in which small group discussion was used, many of the small groups did not use all of the time provided. Therefore, the total number of support episodes to which an individual student could have been exposed was often quite a bit greater in the whole class discussions.

Results such as these were developed for all four matched sets (8 class sections) of the Gravitational Potential Energy lesson sequence. Similar methods were used to analyze another lesson sequence in a related study (Stephens, 2012).

Appendix D. Clustering of Codes

The timeline in Figure D.1, running from left to right, represents 43 min of classroom videotape, about 32 min of which was taken up by small group work, as indicated by the bronze bars on the third and fourth rows. The bronze blocks on the third row show that the teacher stopped by this group 5 times. (Fig. D1 is in color in the web version of this article.)

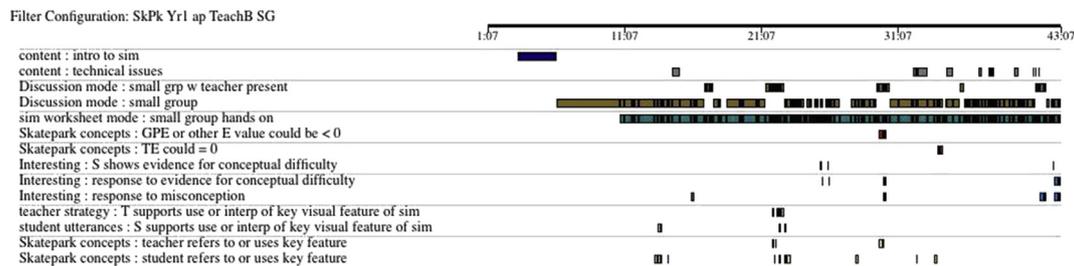


Fig. D1. Clustering of analytical codes around teacher's visits to the small group.

Note clustering of analytical codes (rows 6–14) during and just after the second, third, and fifth teacher visits. In particular, codes in the last 4 rows, for support and use of key visual features, are concentrated around the second teacher visit—including codes for student actions. Additional examples are in Stephens (2012).

References

- Adams, W. K., Reid, S., LeMaster, R., McKagan, S., Perkins, K., Dubson, M., et al. (2008a). A study of educational simulations Part I - engagement and learning. *Journal of Interactive Learning Research*, 19(3), 397–419.
- Adams, W. K., Reid, S., LeMaster, R., McKagan, S., Perkins, K., Dubson, M., et al. (2008b). A study of educational simulations Part II - Interface design. *Journal of Interactive Learning Research*, 19(4), 551–577.
- Bell, R., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction: assessing the inquiry level of classroom activities. *The Science Teacher*, 72(7), 30–33.
- Bennett, J., Hogarth, S., Lubben, F., Campbell, B., & Robinson, A. (2010). Talking science: the research evidence on the use of small group discussions in science teaching. *International Journal of Science Education*, 32(1), 69–95.
- Buckley, B. C. (2000). Interactive multimedia and model-based learning in biology. *International Journal of Science Education*, 22(9), 895–935.
- Chang, H. Y., & Linn, M. C. (2013). Scaffolding learning from molecular visualizations. *Journal of Research in Science Teaching*, 50(7), 858–886.
- Chin, C. (2006). Classroom interaction in science: teacher questioning and feedback to students' responses. *International Journal of Science Education*, 28(11), 1315–1346.
- Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. New York: Springer.
- Clement, J., & Rea-Ramirez, M. A. (Eds.). (2008). *Model based learning and instruction in science*. Dordrecht: Springer.
- Cook, M. (2006). Visual representations in science education: the influence of prior knowledge and Cognitive Load Theory on instructional design principles. *Science Education*, 90, 1073–1091. <http://dx.doi.org/10.1002/sce>.
- Darden, L. (1991). *Theory change in science: Strategies from Mendelian genetics*. New York: Oxford UP.
- Duckworth, E., Easley, J., Hawkins, D., & Henriques, A. (1990). *Science education: A minds-on approach for the elementary years*. L. Erlbaum Associates.
- Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: mental models of electricity. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 99–129). Hillsdale, NJ: Erlbaum.
- Hammer, D. (1995). Student inquiry in a physics class discussion. *Cognition and Instruction*, 13(3), 401–430.
- Hammer, D., Elby, A., Scherr, R., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–119). Greenwich, CT: Information Age Publishing.
- Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, 55(5), 440–454.
- Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan, & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues*. Cambridge, MA: Brookline Books.
- Hutchison, P., & Hammer, D. (2010). Attending to student epistemological framing in a science classroom. *Science Education*, 94(3), 506–524.
- Jones, L., Jordan, K., & Stillings, N. (2001). Molecular visualization in science education. In *Report from the Molecular Visualization in Science Education Workshop*. NCSA Access Center, Arlington, VA, January 12–14.
- de Jong, T., & van Joolingen, W. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179–201.
- Krajcik, J., McNeill, K. L., & Reiser, B. J. (2008). Learning-goals-driven design model: developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education*, 92(1), 1–32.
- Linn, M. (2003). Technology and science education: starting points, research programs, and trends. *International Journal of Science Education*, 25(6), 727–758.
- Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction*, 13, 157–176.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87–99.
- McNeill, K. L., & Pimentel, D. S. (2009). *Scientific discourse in three urban classrooms: The role of the teacher in engaging students in argumentation*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Minstrell, J., & Kraus, P. (2005). Guided inquiry in the science classroom. In *How people learn: History, mathematics, and science in the classroom* (pp. 475–512).
- Nunes-Oviedo, M. C., & Clement, J. (2008). A competition strategy and other discussion modes for developing mental models in large group discussion. In J. Clement, & M. A. Rea-Ramirez (Eds.), *Model based learning and instruction in science* (pp. 117–138). Dordrecht: Springer.
- Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., et al. (2006). PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher*, 44, 18–23.
- Pope, D. C. (2001). *"Doing School": How we are creating a generation of stressed-out, materialistic, and miseducated students*. New Haven, CT: Yale University Press.
- Raghavan, K., Sartoris, M., & Glaser, R. (1998). Why does it go up? The impact of the MARS curriculum as revealed through changes in student explanations of a helium balloon. *Journal of Research in Science Teaching*, 35(5), 547–567.
- Redish, E. (2004). A theoretical framework for physics education research: modeling student thinking. In E. F. Redish, & M. Vicentini (Eds.), *Proceedings of the Enrico Fermi Summer School Course*, CLVI (pp. 1–63). Bologna, Italy: Italian Physical Society.
- Redish, E. F. (2014). Oersted Lecture 2013: how should we think about how our students think? *American Journal of Physics*, 82(6), 537–551.
- Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, 19, 9–20.
- Rieber, L. (1990). Using computer animated graphics in science instruction with children. *Journal of Educational Psychology*, 82(1), 135–140.
- Rosenberg, S., Hammer, D., & Phelan, J. (2006). Multiple epistemological coherences in an eighth-grade discussion of the rock cycle. *Journal of the Learning Sciences*, 15(2), 261–292.
- van de Sande, C., & Greeno, J. G. (2012). *Achieving alignment of perspectival framings in problem-solving discourse*.
- Scherr, R., & Hammer, D. (2009). Student behavior and epistemological framing: examples from collaborative active-learning activities in physics. *Cognition and Instruction*, 27(2), 147–174.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., et al. (2009). Developing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654.
- Smetana, L., & Bell, R. (2009). *Incorporation of computer simulations in whole-class vs. small-group settings*. Paper presented at the NARST 2009 Annual Meeting, Garden Grove, CA.
- Stephens, L. (2012). *Student recognition of visual affordances: supporting use of physics simulations in whole class and small group settings* (Doctoral dissertation). Amherst: University of Massachusetts http://scholarworks.umass.edu/open_access_dissertations/668. ProQuest Dissertations and Theses, 424 p. Open Access Dissertations. Paper 668.
- Tversky, B., Morrisson, J., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4), 247–262.

- Williams, G., & Clement, C. (2015). Identifying multiple levels of discussion-based teaching strategies for constructing scientific models. *International Journal of Science Education*, 37(1), 82–107.
- Williams, M., Linn, C., Ammon, P., & Gearhart, M. (2004). Learning to teach inquiry science in a technology-based environment: a case study. *Journal of Science Education and Technology*, 13(2), 189–206.
- Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: the roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching*, 32(2), 145–160.
- Windschitl, M., Thompson, J., Braaten, M., Kang, H., & Stroupe, D. (n.d.). *A primer on productive classroom conversations*. http://tools4teachingscience.org/tools/discourse_tools/primer.html Accessed 13.12.14.
- Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, 96(5), 878–903.
- Woods, D., & Fassnacht, C. (2007). *Transana v2.30b*. Madison, WI: The Board of Regents of the University of Wisconsin System. <http://www.transana.org>.
- Wu, H.-K., & Huang, Y.-L. (2007). Ninth grade student engagement in teacher-centered and student-centered technology-enhanced learning environments. *Science Education*, 91, 727–749.
- van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *Journal of the Learning Sciences*, 6(2), 227–269.
- Zietsman, A. I., & Hewson, P. W. (1986). Effect of instruction using microcomputer simulations and conceptual change strategies on science learning. *Journal of Research in Science Teaching*, 23(1), 27–39.