DESIGNING CLASSROOM THOUGHT EXPERIMENTS: WHAT WE CAN LEARN FROM IMAGERY INDICATORS AND EXPERT PROTOCOLS[†]

This study explores roles that thought experiments (TEs) play in the classroom. An in-depth analysis of case studies identifies multiple instances of TEs in classroom episodes. With the use of a detailed list of imagery indicators, evidence is provided for the involvement of kinematic and kinesthetic imagery. Student- and teacher-generated TEs appear to be an important form of reasoning in the science classrooms investigated. There appear to be important similarities between expert and student uses of TEs, and evidence supports the hypothesis that certain TEs can allow students to evaluate the consistency of newly constructed models. Students are observed to reason with each other's TEs and those generated by the teacher as an important part of classroom discussion, and they appear to have considerable value as a sense-making strategy.

A. Lynn Stephens, University of Massachusetts, USA
John J. Clement, University of Massachusetts, USA

Objectives

As part of an ongoing concern about the ability of science students to revise and evaluate their mental models, there has been interest in students' ability to generate and use thought experiments (TEs). We are particularly interested in whether TEs can be documented in large class discussion. In the present study, the existence and possible importance of thought experiments in the classroom are explored by coding transcripts of classroom discourse. We investigate whether TEs are associated with evidence for the use of imagery, whether we see evidence that students actually run the TEs proposed by the teacher, whether students generate their own TEs, and whether students can evaluate the TEs of others. Using a detailed list of imagery indicators derived from expert protocols, we conduct in-depth analyses of a short episode that occurred in a physics classroom in an urban high school in the Northeastern US, and a much longer episode from a college preparatory physics class using an innovative curriculum (Camp, Clement, et al), located in a middle class suburban high school. An important goal is to reveal factors that can make classroom TE's effective.

[†] This material is based upon work supported by the National Science Foundation under Grant REC-0231808, John J. Clement, PI. 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.

Theoretical Framework

There is previous research on the structure of TEs (Brown 1986); the use of imagery in problem-solving in physics (Clement 1994); student inquiry in science (Driver 1983); and the function of TEs in scientific thinking (Kuhn, 1977). Giere (1988) and Darden (1991) argue that the ability to generate and evaluate mental models is a crucial aspect of science, and Nersessian (1992, 1993) and Gooding (1992b) believe that one powerful way to do this is to run a thought experiment. There has been much recent research in students' use of visual and kinesthetic imagery (Hegarty, 1992; Kozhnevnikov, Hegarty & Mayer, 1999) and the importance of mental modeling (Gentner, 2002; Giere, 1988; Nunez-Oviedo, Rea-Ramirez & Clement, to appear). Recent research on TEs has investigated students' ability to generate and use TEs in small-group collaborative settings (Reiner, 1998; Reiner & Gilbert, 2000); the value of divergent, qualitative thinking methods including new knowledge obtained through TE's (Gooding, 1992a, 1992b, 1996); the nature and function of TEs in scientific discovery (Clement, in press); and the importance of TEs in teaching and learning (Clement & Steinberg, 2002; Gilbert & Reiner, 2000). However, very few of these studies investigate the role of TEs in full class discussion. TEs have more often been studied within the context of small-group sessions or individual think-aloud protocols (Clement, in press). Hammer (1995) described thought experiments in physics class discussions as one of several kinds of process skills that were exhibited by students when the teacher in his case study took care to foster an open attitude toward contributing ideas. Nunez-Oviedo (2003) has analyzed the role of TEs in the classroom from the standpoint of the apparent uses to which they are put by the teacher, but we believe there is much more that can be learned about the role of TEs in the classroom, and especially of the uses students make of them. Specifically, there has been little investigation into the roles TEs may play in supporting model construction in large class discussion. There has also been little, if any, analysis of the possible association of student use of TEs with student use of imagery in large class discussion.

Significance

Before addressing the question of whether students can run TE's, we first must clearly define what we mean by a TE. A number of authors have speculated that TE's make extensive use of mental imagery (Sorensen, 1992) and that both dynamic imagery (with kinematic and/or kinesthetic components) and embodiment play a crucial role (Gooding, 1992b; Nersessian, 1992; Reiner, 1998; Reiner & Burko, 2003; Reiner & Gilbert, 2000). It is generally assumed by these authors that a TE is an experiment carried out in thought and that an outcome must be predicted or inferred. However, even though a typology of TE's has been developed (Reiner & Burko, 2003) and different stages of TE's identified (Reiner, 1998), the definitions provided have not included observable behaviors, making it difficult to investigate the possible occurrence of TEs in the classroom. This issue is addressed here by adopting the following definition (Clement, in press):

> *Performing an (untested) thought experiment (in the broad sense):* the act of considering an untested, concrete system* (the "experiment" or case)

> and attempting to predict aspects of (or underlying causes for) its behavior. Those aspects of behavior must be new and untested in the sense that the subject has not observed them before nor been informed about them.
>
> > **Concrete system* here means one involving concrete objects or experiences (and relationships between them) rather than involving abstract higher order relations only.

Previous authors have not made it clear whether it is necessary that the purpose of the experiment be to test a concept, model, or theory; however, at times this has been strongly implied (Gilbert & Reiner, 2000). Elsewhere, one of us (Clement, in press) has considered the addition of this requirement to result in a definition of TEs in a narrower sense:

> *Performing an evaluative Gedanken experiment (in the narrow sense):* the act of considering an untested, observable system designed to help evaluate a scientific concept, model, or theory—and attempting to predict aspects of its behavior.

Although we do not go into depth about the distinction between the two definitions here, we do note when a TE under discussion appears also to fit the above definition of a Gedanken experiment. Unless stated otherwise, it may be assumed that we are referring to TEs in the broad sense.

Whether TEs are considered in the broad or the narrower sense, there is some evidence that they can involve imagery-rich mental simulation and that this dynamic imagery can enable the user to access implicit knowledge, making it more explicit (Clement, 1994, to appear). There is also evidence that some students can and will use TEs to find solutions to problems if the problems are formulated in a way to encourage this kind of solution process (Reiner & Gilbert, 2000). Reiner & Gilbert believe that only a small portion of the implicit knowledge that students access through TEs can be articulated verbally. By using methods originally developed in the in-depth analysis of expert protocols (Clement, Zietsman and Monaghan, 2005; Clement, 1994), identification of imagery-use indicators has allowed us to further address the question of whether classroom TEs can involve dynamic imagery. Our analysis of the teaching tapes leads us to believe that these properties may be able to be fostered instructionally.

Design and Procedure

A number of transcripts of classroom activity have been previously examined; in most of this classroom activity, inquiry-based methods of teaching and learning were being employed. TEs were tentatively identified and categorized as student generated or teacher generated. Portions of two transcripts have been selected for more detailed analysis and coded for the presence of TEs. This has been done by breaking the definition for TE in the broad sense into two requirements, which we have coded for separately:

1. Subject attempts to predict behavior of concrete system;
2. Subject has not observed the experiment before, nor been informed about its behavior;

For example, in one of the transcripts we examine, a student refers to a chalkboard drawing of the Earth with a stick figure of a man standing on it (as in Fig. 3):

> S5: Well I just think that gravity has nothing to do with rotation, but maybe with ***[[quick rotating movement with right forefinger]]*** rotation like ***[[points to chalkboard]]*** that guy is trying to get ***[[emphatic movement with his right hand and arm, beginning on the right side of his body and sweeping leftward in front of him]]*** thrown off the Earth. So he's getting ***[[repeats sweeping movement]]*** pulled at the same rate but he's also getting ***[[reverses previous movement, pulling his right hand and arm back to the right]]*** pushed away.

S5 attempts to predict the behavior of a concrete system, that of a rapidly rotating Earth with a man standing on it. His gestures seem to depict an object in the air immediately in front of him. He has never observed the Earth from this vantage point, and certainly has not observed it spinning rapidly enough to throw a person off. We categorized the above incident as a student-run TE. (This will be discussed more in a later section.)

In another example from the same transcript, the students are looking at a bell jar that is in the process of having its air evacuated by a pump. The question arises as to how they will know when the pump is done, and one of the students offers that they will know when it is really done when the jar implodes. The teacher asks the students what they think of this.

> T: I'm asking a hypothetical question. If we can get a perfect vacuum of air is it possible to make glass strong enough to withstand the air pushing it.
>
> S5: Sure.
>
> S: Probably.
>
> S: No.

Here again, the students are making predictions for a concrete system that was proposed by the teacher. However, since there are other pieces of physics demonstration apparatus that use evacuated glass containers, there is not enough information in their answers to make a reasonable determination as to whether the system is untested for them, so we did not categorize these as TEs.

Once episodes were coded as having evidence for TEs, they were then analyzed for the following factors:

- whether the TE was generated by the teacher or the student,
- whether the TE was run by the teacher or the student,
- the problem that was targeted by the episode,
- whether the intent of the experiment fits the broad definition of a TE or the more restricted definition of a Gedanken experiment;
- whether an analogy or an extreme case are present.

The distinction between generating a TE and running it is an important one. A pedagogical TE may be generated in order to ask one's audience to make a prediction about a system where the results are unknown to the audience but known to the generator. Often, the pedagogical TEs in the transcripts analyzed here were generated by the teacher and run by the students; however, there are several incidences where we believe a student generated a pedagogical TE in order to convince fellow students of a point.

What follows is a subset of a much larger set of indicators developed by Clement, et al. (2005).

---

1. <u>Imagery reports</u>: Subject states that s/he is imaging or imagining, "seeing" or "feeling."
2. <u>References to perceptions</u>: Similar to imagery reports but not as direct. The subject refers explicitly to the sensation of perception while describing visual or other perceptual aspects of scene during thinking by using phrases such as "the car probably looks like it is going that way".
3. <u>Depictive Motions</u>: Subject makes non-stylized hand or body motions depicting objects, forces, locations, or movements of entities.
4. <u>Motions Depict Dynamics</u>: Hand motions depict the form of a dynamic event, not simply a static picture.
5. <u>Dynamic Imagery Report</u>: An imagery report where the subject indicates that they are imagining motions, changes, or interactions over time in a situation.
6. <u>Personal Movement Projection or Analogy</u>: (a) Refers to movements of entities in target situation as if they were moved by a person, or (b) uses a personal analogy by referring to an analogous situation involving the body.

7. Uses Force Terms: Imagery indicator occurs in conjunction with force context terms such as "pull", "push", "twist", "effort", "try to move it."

8. Focuses on a drawing while making statements about spatial or physical relationships between entities in the drawing and other entities or actions not depicted in the drawing.

Table 1. Imagery Indicators from Clement, et al. (2005).

It should be noted that items 4, 5 and 7 are not imagery indicators on their own, but complement the others by pointing to a particular type of imagery. Items 4 and 5 indicate that the imagery is dynamic rather than static, and 7 indicates that the imagery is kinesthetic.

The pool of videotapes on which we drew for this study comes from a number of settings where inquiry-based methods of teaching and learning have been employed. As an initial exploratory study, we have analyzed 4 sessions in which TE's were generated and run by experts and 6 sessions involving TE's from pedagogical contexts. Here, we report on a short episode that occurred in a physics classroom in an urban high school with mixed ethnic backgrounds and a much longer episode from a physics classroom in a suburban high school.

Evidence From A Classroom: The Waterwheel.

Our first study is a short episode and will allow us to explain the methodology.

The school is located in an urban area in the northeastern US and has a mixed ethnic background. The students were in a classroom on an upper floor of the building, with a window that looked down over the street.

The lesson was an inquiry into variables involved in building waterwheels (WGBH Videos, no date). Groups of eighth grade students, working in teams for six weeks, were asked to design and construct waterwheels out of everyday materials provided; the waterwheel had to be able to lift a cup of nails. Then, the students were challenged to try to maximize some effect. For example, they could alter their waterwheel in order to investigate how to lift the most nails or how to lift the same amount of nails the most quickly. After each project, the various groups reported their results in large class discussion. The dialog we analyze occurred during one of those reports.

The Transcript

The twenty-four lines of transcript are included as Appendix A. The teacher describes how one of the groups had experimented with the placement of the hose that was used to shoot water onto their water wheel. The students had tried holding the bucket and hose up high so that the water would fall with greater force. However, S1 did not think he was being very accurate in aiming the water from this height and asked the teacher for a longer hose, which the teacher apparently did not have. Now, the teacher asks the class to consider a hypothetical situation where they do have a longer hose, and requests that they predict the effect. (See Figure 1- in both the case of the long and short hose, the water starts from the same height and falls the same distance to the wheel, and they assume that the water will hit the wheel).

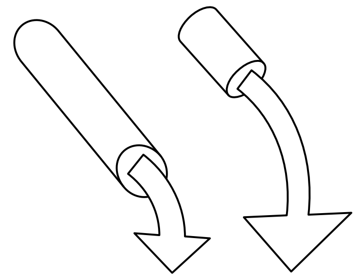

Figure 1: Hoses of different lengths

The teacher's request and the next several lines of dialog are shown below. Our initial interest was in the passage that follows this one. However, when applying the definitions above, unanticipated evidence for TEs was revealed:

I:	Do you think, do you think if you still had a longer hose, if it were traveling through the hose at that distance that you would have gotten the same results?	TE: pedagogical Requests prediction of untested, concrete system
S3	Oh! No! OOOhh OOO OOO!	
S1:	[Probably not.]	
S1:	No, it wouldn't (S).	
S2:	I think because. I think the gravity would have had the same pull on it... even though it was out in the air. It has that gravity, that pull. And I think that even though if it is in the tube I think the pull is the same.	Runs TE Predicts force would be the same.
I:	(S4)?	
S4:	I think the force would have been greater. It is sort of like a roller coaster when you go down the force is a lot greater when you go down than when you go up. So like the longer, if it's longer, right, it's faster... like it will go faster and the force will be greater.	Runs TE Predicts force would be greater.

Immediately following this passage, S1 generates a thought experiment and then modifies it in a way that makes it more powerful. We coded this section for the presence of depictive gestures and other imagery indicators. Here is about a third of the imagery coding, a single sentence of S1's dialog.

S1	I think the force, if the **[[1]]** line is shorter, it would be more.	3. Depictive motion
	1. Holds hands up in front of himself close together and moves them apart, left hand ending high and right hand ending low, as though starting in the middle of a diagonal line and tracing it to its ends.	
	Cause like say, say for example you're going down a slide,	<u>6a. Personal mov proj</u>
	right, and if you, if you are going out a **[[2]]** tall slide that is	<u>3. Depictive motion</u>
	2. Holds left hand up higher than his head, fingers flat and pointing downward at a steep angle.	
	going off like, **[[3]]** Lincoln School or something like that, it could	
	3. Holds right hand far up above his head and points in the direction of the window, but with pointer finger angling down. Looks in direction he is pointing.	not depictive
	be like **[[4]]** real steep like this or something like that, you'll	3. Depictive motion
	4. Holds right hand up with fingers flat and pointing down at an angle, as though indicating a slide starting about head-high on his right and angling down toward his left.	
	[[5]] go slower, because you're, you are **[[6]] hitting** the slide and the	<u>4. Motion depicts dynamics. 14. Uses force term w/motion</u>
	5. Moves right hand downward twice, as though along the slide. 6. Moves both hands up to his shoulders and gestures toward them with a quick flick.	
	slide is slowing you down so it's like the **[[6]]** friction. Right?	4. Motion depicts dynam.
	6. Waves hands slightly in opposition to each other, fingers extended as though brushing the air	

The results of the coding are discussed in the following section.

<u>Coding for TEs</u>

The three cases in this episode are shown in Figure 2.

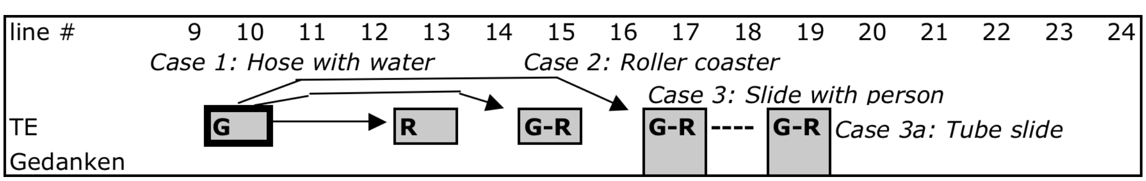

Figure 2: The three Waterwheel cases

In Line 10, the teacher generated the case that paired water flowing through a hypothetical long hose with water flowing along the same path through the shorter hoses that they had. We have represented this with a G for "generated" and a heavy border to indicate that it was by the teacher.

Next, a student ran the case as a thought experiment; that is, a student made a prediction that apparently did not result from simply replaying a previously observed event, nor from repeating information gained from an outside authority. The student predicted that the force of the water would not change and gave a logical reason for that prediction, saying that gravity is the same inside as outside the hose. He evidently ran the experiment proposed by the teacher in order to generate this prediction.

In Line 15, a student made the prediction that water would emerge from the longer hose with more force because it would be faster. She then made an analogy with a rollercoaster ride, reasoning that if the ride were longer, you would go faster. In this case, she has generated a new case and apparently run that case as a TE, making her prediction by analogy. Her analogy is faulty. However, this does represent an interesting attempt at reasoning that goes beyond quoting a textbook.

In Lines 17 and 19, the student who had asked for the longer hose also reasoned by analogy in a TE. He compared water flowing down the hose to a person sliding down a slide, reasoning that a person going out of the upper story window of their classroom would "be bust up more" if they just fell than if they went down a slide. In the face of incredulity from his classmates, he then asked them to imagine the slide wrapped all around themselves:

> S1: Like a tube slide, it's going straight down. When you're going down it, like your arms is hitting the sides of the tubes and stuff like that, and keep on bumping into it. Just like the water bumps into the sides of this tube right? So it will slow you down because of friction, right? and- but if you just drop, right? the only friction you'll be getting will be from the air and the air wouldn't cause as much as from the tube because the tube is a solid and the air's a gas.

He not only generated the TE and ran it, but then improved it and ran it again. He used this case not only to generate a prediction by analogy, but to evaluate the cause of the behavior he was predicting. As his words seem to indicate that he was evaluating friction as a new component of his mental model of how water flows through a hose, we have coded this incident a TE in the narrow sense, an Evaluative Gedanken Experiment.

<u>Coding for Imagery</u>

In the three lines of S1's TE, he is observed making 16 depictive gestures. These are not the only gestures he made, as he frequently pointed or moved his hands for emphasis. But there were 16 incidents where he appeared to be creating a picture of the tube in the air, and almost none of these gestures were repeats. He indicated the position of the slide, the angle of the slide, the force of the water against the slide, the slide growing into a tube around him, his forearms bumping against the slide, his chest brushing the slide, etc.

The first depictive gesture in Line 17 is paired with a Personal Movement Projection,

> S1: Cause like say, say for example you're going down a slide, right, and if you, if you are going out a ***[[G]]*** tall slide. . .

where the student is substituting himself for the water. These words overlap his forming of the gesture, which depicts the shape and angle of the slide. Later in the same line, he says

> S1: You are ***[[G]]*** hitting the slide. . .

while moving his hands in quick flicks toward his shoulders [6]. Rather than pointing to the places on his shoulders that would get hit, he appears to be demonstrating the actual hitting. "Hitting" is a force term, and, since it is used in conjunction with a gesture that appears to exemplify it, it seems reasonable to hypothesize that S1 is using mental imagery of a kinesthetic nature, imagining his body getting hit by the slide as he goes down. In Line 21, there are two other depictive gestures that occur in conjunction with the use of force terms that appear to describe the gestures.

To summarize, there is evidence here that three students in this classroom were able to run TEs. For at least one of the TEs, there is triangulated evidence for the involvement of imagery when the student evaluated the case he had created. The prior knowledge that was tapped during the running of this experiment appeared to be kinesthetic in nature, as the student applied a schema for anticipating the effects of friction on his body to generate a prediction for how water would behave in a hose.

Clement (to appear) has noted the importance in expert thinking of experts' ability to improve their thought experiments to increase confidence in the results. In the case of student S1, there is evidence that he was able to do exactly that.

Evidence From a Classroom: Gravity

This large class discussion occurred over the span of two days in a senior level high school physics class in a middle class community in the Northeast. The discussion was videotaped as part of the development process of a physics curriculum (Camp and Clement, et al., xxx). The class had a total of 15 students, four of whom were absent on the first day. The transcript begins when the topic of gravity was introduced by the teacher and comprises 42 minutes of tape, covering the last part of one class period and first half of another.

<u>The Transcript</u>

We have identified six separate cases (concrete examples of systems) that are topics of discussion. Many of these cases are incorporated into TEs by the students, some of then into more than one TE. Two of the cases are introduced by the teacher as part of the planned lesson, but these subsequently undergo numerous modifications by the students. Four cases are introduced by students.

The discussion begins with the teacher asking the students to consider a drawing on the board, recreated in Figure 3.

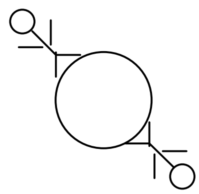

Figure 3. US/Australia Case

The teacher explains that the upper stick figure is standing in the US and the lower in Australia.

> T: Now. Vote Number One—try to listen so I don't have to repeat it twice—ah, compared to the United States, gravity in Australia is: a little less, equal, a little bit more.

Students have differences of opinion on this, leading to a very active discussion. (For an investigation into the design characteristics of this teacher-designed case, see Stephens & Clement, 2006, this volume.)

The six cases are summarized in Table 3.

Case 1: Standing on a round object	Teacher introduced
Case 2: People and small objects pulling on each other	Student introduced
Case 3: Vacuum chamber and effect on weight	Teacher introduced
Case 4: Spinning carnival ride	Student introduced
Case 5: Catapult	Student introduced
Case 6: Spinning spaceship	Student introduced

Table 2. The Six Cases

This transcript was coded in entirety for TEs. In this very early stage of the study, a major focus is on testing and strengthening the wording of the definitions. The task is to apply definitions derived from expert cases to the more scanty evidence available in large class discussion. Hand motions are evident and predictions are abundant in this videotape. However, students frequently give little information about how they have arrived at their predictions. At times it is not possible to tell whether a student is repeating a correct answer s/he has heard elsewhere, is more interested in getting a reaction from classmates than in obtaining a useful answer, or has actually run a thought experiment but done so poorly, perhaps failing to establish the appropriateness of an analogy. Much work has been required to hone our definitions to give them power to say yea or nay in these borderline cases. Nonetheless, there have been many cases on which we have agreed from the outset, and it is primarily on these cases that our discussion rests.

In contrast, the imagery indicators for depictive hand motions have proven fairly easy to deploy, and the entire transcript has been coded for hand motions.

After reaching agreement on the coding for the hand motions, the verbal imagery indicators, and the TEs, we compared the results to see how often imagery indicators coincided with evidence for TEs. Those results are presented below.

<u>Coding for TEs and Imagery</u>

Figure 4 is a chart of the results of the coding for TEs and imagery use for the Day 1 transcript. It is the same kind of chart we used for the Waterwheels cases, though more complex.

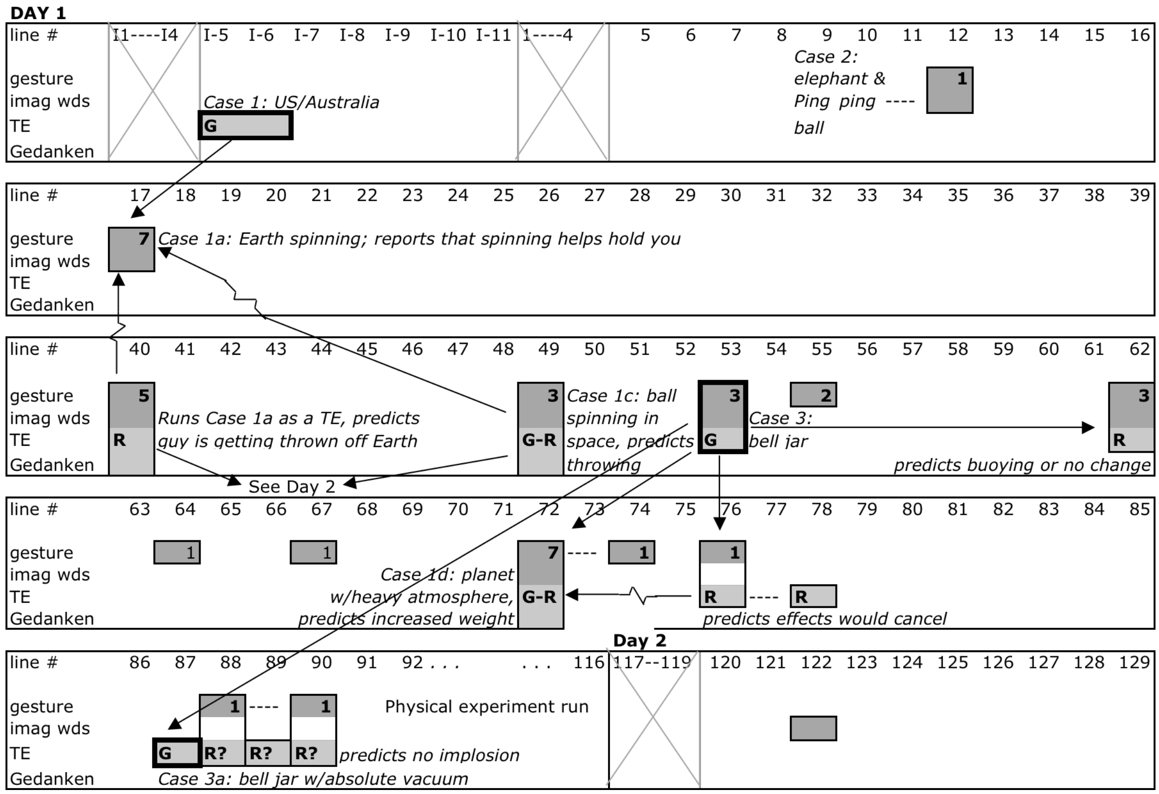

Figure 4. Gravity class TEs and imagery use, Day 1.

Key

- *Either gestures or verbal imagery indicators are present.*
- *Both gestures and verbal imagery indicators are present (triangulated evidence).*
- *There is evidence for a TE in the broad sense.*
- *There is evidence for a TE in the narrow sense, an Evaluative Gedanken Experiment.*
- *The teacher introduced the case.*
- → *The later case is a variation of the earlier case or incorporates it.*
- *The later case appears designed to dispute the results of the earlier one.*
- 7 — *7 depictive gestures were coded for this line of dialog.*
- G-R — *There is evidence that a TE was Generated and Run.*

In transcript lines I-5 and I-6, the teacher introduces the US/Australia case. This case asks the students to predict whether or not there will be a change in weight for a person who moves from the one location to the other. A student mentions the spinning of the Earth, which introduces it into the discussion. We count the Spinning Earth as Case 1a. Imagery indicators, both gestural and verbal, are present. However, the student's statement "spinning holds you" does not appear to be a prediction, but rather a report of the student's belief. Therefore, his statement is not taken as evidence for a TE.

In Line 40, another student responds to the comment of the first when he says,

> S5: Well, I just think that gravity has nothing to do with rotation, but maybe with rotation, like that guy is trying to get thrown off the Earth. So he's getting pulled at the same rate but he's also getting pushed away.

This student makes a prediction that spinning will throw "that guy," which becomes a hot topic of debate in the class. (Note that he speaks of "that guy"as though it were the drawing on the board with its stick figure that is doing the rotating.) Although we cannot tell for sure that the spinning system is untested for this particular student, he has worded his reply as though he is thinking through the spinning-as-throwing idea as he speaks. The student appears to use the case to help evaluate the effect of spinning in his mental model of gravity, so we have classified it as a TE in the narrow sense, a Gedanken experiment.

Also evident during the two sentences of Line 40 are no less than five depictive gestures. (See the Design and Procedure Section, above, for Line 40 with gestures included.) These occur alongside force terms (thrown, pulled, pushed) and appear to depict them. This is one of a number of cases where the imagery indicators occur in close proximity to the running of a thought experiment.

The transcript of the first day provides sufficient evidence to code five incidents as involving the generation of thought experiments, two of them by students. Both student-generated TEs are accompanied by gestures. In one of them, the hand motions are paired with force terms, "Air pressure ***[[G]]*** pushes on everything." The other incident involves an abundance of imagery indicators, including a dynamic imagery report, "Let's just imagine a ball floating in space you tape your feet to. . . And you start spinning the ball around;" a kinesthetic imagery report, "you're gonna feel like you're gonna be thrown off;" and a personal force analogy, "you're just gonna feel the forces being spun around."

In addition, there is evidence that two students ran TEs generated by the teacher. These were both accompanied by gestures. In one of them, the student focuses on a drawing of a bell jar on the board and moves his hands as he talks about the possibility of the air buoying up the weight that is depicted inside the jar. At other points, students appear to be running mental simulations and generating predictions, but in each of those cases there is not enough information to determine whether the system in question was untested for those students. Coding in this conservative manner yielded five incidents in less than 20 minutes of tape where there was evidence for the generation or running of TEs paired with multiple indications of imagery use. Four of these were student incidents, involving three separate students.

Day 1 ended with a physical experiment with a bell jar to test predictions the students had made for Case 3, a teacher generated TE concerning weight in a vacuum.

The results of coding for the second day are below.

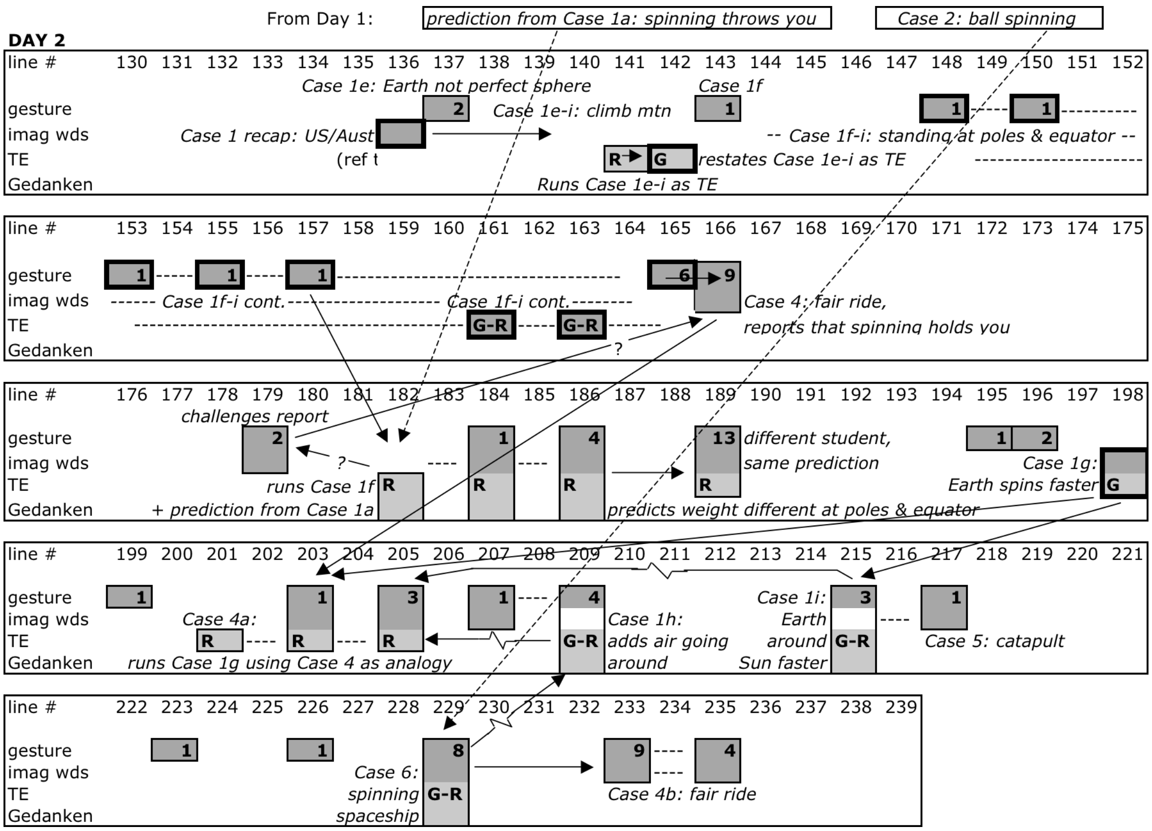

Figure 5. Gravity class TEs and imagery use, Day 2.

This class began with the students energetically talking about questions raised the day before. A few minutes into the discussion, the teacher repeated the US/Australia case for the benefit of the four students who had been absent. This was followed by a new round of discussion in which, over the next 25 minutes, there is evidence for the generation of six new thought experiments, the first three by the teacher and the last three by students. In addition, there were four incidents involving three students where the students appeared to run thought experiments generated by other students or by the teacher. Of these ten incidents, seven occurred along with one or more imagery indicators. Four of these had multiple imagery indicators, and in each case, they provided evidence for the use of kinesthetic imagery; the students used force terms and hand motions that appeared to depict those forces. Perhaps this is not surprising, as the topic of all of these was whether or not one would be thrown off a spinning object. It is noteworthy, however, that the students appeared to be actively reasoning.

One thing that struck us as we reviewed these transcripts was that there is evidence of at least two students adopting, evaluating, and improving upon the imagistic cases proposed by other students. On Day 2 (Lines 182-186), S4 tries repeatedly to argue that the spinning of the earth would have an effect on your weight. However, he appears to have *weight* conflated with *gravity* and so has trouble understanding what S5 is trying to tell him; i.e., that changing your weight will not affect gravity. Finally, after nine turns of speaking between S4 and S5, S9 is called on. He appears to adopt the imagery of S4:

> S4: OK, say that it's [the spinning Earth] throwing you. Then that still means that the top [the North Pole] is still gonna be throwing you left and at the side [the equator] *(indistinguishable)*. So your weight's gonna be different.
>
> S5: Yeah, but that has nothing to do with gravity.
>
> S4: Why not? What if the earth is trying to throw you around at the equator?

This back and forth continues between the two classmates for three more turns. Then the teacher calls on S9.

> S9: What were we arguing about? Well, I'm basically taking (S4's) position in that when the Earth spins, it seems logical to me, although (S10) says its wrong, but it seems logical to me that there would be a force...say you're on the equator and you're going around, there's this greater force pushing you off the Earth than if you were on the pole and you're doing this little circle. It's just much less of a force throwing you that way. But if gravity is the same ***[[G]]*** here, and gravity is the same ***[[G]]*** here, it seems that you would weigh less here because you're being thrown off more that way. Although you'd still stick to the Earth. You could still...I think you would weigh less.

S9 has his terms more clearly distinguished, but he also is able to describe S4's TE in a more clearly worded manner.

In a separate incident on Day 1, a student appeared to adopt another student's imagery, convert it into an extreme case, and then run it as a Gedanken experiment. In Line 17, S4 said he thought "somehow the fact that [the Earth] spins causes a lot of the main force of gravity." Several students had attempted to address this misconception--apparently without success, because both S4 and S6 continued to defend rotation as a cause of gravity. A few minutes later, S7, who had been quiet until this point, suggests the following.

> S7: Well, in reference to rotation and gravitational force, I think of them as being two opposite forces because if you stand on...let's just imagine a ball floating in space you tape your feet to. And you start spinning the ball

> around, you're gonna ***[[G]]*** feel like you're gonna be ***[[G]]*** thrown off. But if it's a small ball, then the attraction between you and that little small mass is negligible so that you're just gonna ***[[G]]*** feel the forces being spun around in a centrifugal force.

The massive earth has shrunk to a small ball and the spinning has increased from one revolution a day to many times a minute. This extreme case also occurs with multiple imagery indicators.

It should be noted that another possible interpretation of this experiment, besides being an evaluative Gedanken experiment, is that it was pedagogical in intent. S7 was trying to communicate his model of gravity and rotational forces to his fellow students, as many of these students were, but the question is whether he was already clear about the implications of his experiment before he developed it here. He does not cite an authority and one receives the impression that, if he is not running it as a TE now, he may have considered something similar in the past. (Incidentally, if it was intended pedagogically, it may not have fallen on totally deaf ears; it is possible that this case helped inspire the spinning spaceship case of Day 2. This is reflected in a long dashed arrow in Figure 5.) For purposes of this paper, we have elected to leave this case in the Gedanken category.

In total, the 42 minutes of videotape contains 11 episodes coded as TE generation, 5 of them by students, In addition, there are 7 cases of students apparently running TEs formulated by others, for a total of 18 episodes where we have evidence that TEs are being generated or run, or both.

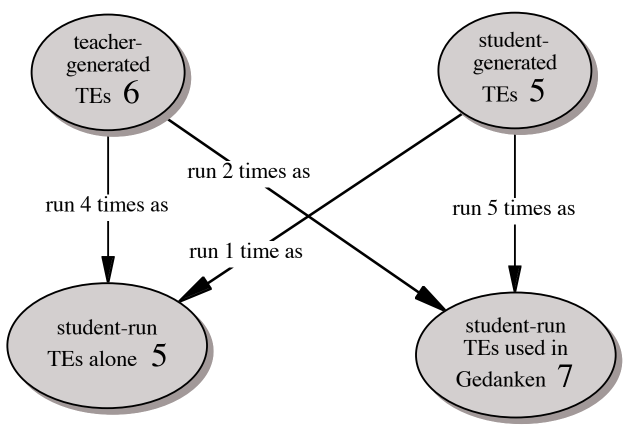

Figure 6. Breakdown of TEs: TEs were often run multiple times and in various combinations, so the arrows and circles do not add up.

Portions of the coded transcript are in Appendix B.

Findings

1) Evidence for thought experiments, as defined above, was found in the episodes. In the Waterwheels episode, there is evidence for one teacher generated and 2 student

generated TEs, with 3 instances of TEs being run. In the Gravity tape, there is evidence for 6 teacher generated and 5 student generated TEs, with 12 instances of TE's being run. This represents only the two classroom episodes that were analyzed here. There is no reason to believe that TEs cannot occur in other large class discussions.

2) There is evidence for the involvement of imagery and imagery-rich mental simulations in the running of the TE's. There were 13 instances where there was evidence for imagery use during the tube slide Gedanken experiment in the Waterwheels tape, and 32 instances in the Gravity tape. Several instances from the Waterwheels tape and half of those from the Gravity tape had triangulated evidence for imagery from multiple indicators. Many of these imagistic episodes were associated with evidence for TEs.

3) There appear to be important similarities between expert and student uses of TE's:

> a. Experts sometimes use TEs as pedagogical devices, in an attempt to communicate their models to others. There was evidence here that some students attempted to do the same.
>
> b. Experts have been observed modifying their TEs to enhance the use of imagery and to clarify the implications (Clement, 2005; Clement, to appear); some of these students appear to do the same. The most notable example was S1 in the Gravity tape, when he modified a flat slide into a tube slide with the result that the potential for friction was maximized.
>
> c. A set of expert strategies for the refinement of imagery has been shown in Stephens & Clement (2006, this volume) to be applicable to student-designed cases. The frequency with which evidence of imagery use was observed in conjunction with evidence of TEs suggests that these strategies could prove helpful in the development of effective TEs.

4) Students can formulate and run evaluative Gedanken experiments. There appear to have been numerous instances in these class sessions when TEs were used by the students to evaluate explanatory models. A few of these were discussed, but, as a look at Figures 4 and 5 will reveal, there were many other instances noted.

5) Evidence was presented that much of the imagery used in these episodes appears to be of a kinesthetic nature. The most frequent form of evidence for imagery use in these transcripts was the use of force terms coupled with gestures. These gestures appeared to depict what the force terms were describing.

TE's evidently can have considerable power to convince, either by exposing inconsistencies in one's conceptions, as they may have done for S4 in the Gravity tape, or by strengthening one's conviction in an outcome, as they appeared to do for S1 in the Waterwheels tape. This argues for their value as a sense-making strategy. (Although we would not suggest that they replace real experiments in the classroom, this suggests that they can be an important complement. In fact, TEs and real experiments seemed to serve complementary roles in both classroom episodes analyzed here.)

We can speculate that TE's in the broad sense may be more widespread in classroom settings than has been commonly thought, especially in situations where students are asked to make predictions for systems they have not yet observed. Our evidence supports Reiner's (1998) speculation that TE's are natural processes in science learning

The widespread use and apparent effectiveness of TEs in a number of classrooms for which we have tapes, of which the two analyzed here are but examples, suggest that it would be of value to educators to understand what goes into fostering an effective classroom TE. To that end, we can note some commonalities in the teacher-generated cases we have studied, without implying that these are the only attributes that are useful. The teachers may have used some of these strategies instinctively, rather than consciously.

- The cases were used in the context of inquiry methods of teaching, where prior conceptions were encouraged to surface and to be debated.
- Many of the teacher- and curriculum-generated cases were intended to target misconceptions that the teachers believed to be prevalent for that topic and age group.
- The cases appealed to both kinesthetic and visual modalities.

Finally, each of the teachers appeared to generate cases intended to foster vivid imagery, and data from imagery indicators provided evidence that the cases succeeded in this.

References

Brown, J. R. (1986) The structure of thought experiments. *International Studies in The Philosophy of Science: The Dubrovnik Papers*, 1-15.

Camp, C., Clement, J., Brown, D., Gonzalez, K., Kudukey, J. Minstrell, J., Schultz, K., Steinberg, M., Veneman, V., and Zietsman, A. (1994). Preconceptions in mechanics: Lessons dealing with conceptual difficulties. Dubuque, Iowa: Kendall Hunt.

Clement, J. (1994). Use of physical intuition and imagistic simulation in expert problem solving. In Tirosh, D. (Eds.), *Implicit and explicit knowledge: An educational approach*. Ablex Publishing Corp.

Clement, J. (2002). Protocol evidence on thought experiments used by experts. Gray, W., and Schunn, C. (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Clement, J. (in press). Thought experiments and imagery in expert protocols. In Lorenzo Magnani, Ed. *Model based reasoning in science and engineering*, London: King's College Publications.

Clement, J. (to appear). *Model construction in scientists and students: Case studies of imagery, analogy, and physical intuition.*

Clement, J. and Steinberg, M. (2002). Step-wise evolution of models of electric circuits: A "learning-aloud" case study. *Journal of the Learning Sciences, 11*, 389-452.

Clement, J., Zietsman, A., Monaghan, J. (2005). Imagery in science learning in students and experts. Gilbert, J. (Ed.) *Visualization in Science Education*. Dordrecht, The Netherlands: Springer.

Darden, Lindley (1991). *Theory change in science: Strategies from Mendelian genetics*. New York: Oxford.

Driver, R. (1983). *The pupil as a scientist?* Milton Keynes: Open University Press.

Gentner, D. (2002). Psychology of mental models. In N. J. Smelser & P. B. Bates (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 9683-9687). Amsterdam: Elsevier Science.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: Chicago University Press.

Gilbert, J. K. & Reiner, M. (2000). Thought experiments in science education: Potential and current realization. *International Journal of Science Education, 22*(3), 265-283.

Gooding, D. (1992a). The procedural turn: Or, why do thought experiments work? In Giere, R. (Ed.) *Cognitive models of science*. Minneapolis: U. of Minnesota Press.

Gooding, D. (1992b). What is *experimental* about thought experiments? In Hull, Forbes, & Okruhlick (eds.), pp. 280-290.

Gooding, D. (1996). Creative rationality: Towards an abductive model of scientific change. *Philosophica, 58* (2), 73-102.

Hammer, D. (1995). Student inquiry in a physics class discussion. *Cognition and Instruction, 13,* 401-430.

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1084–1102.

Kozhevnikov, M., Hegarty, M., Mayer, R. (1999). *Students' use of imagery in solving qualitative problems in kinematics*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Kuhn, T. S. (1977). The function of thought experiments. In T. Kuhn (ed.) *The essential tension*. Chicago: University of Chicago Press.

Nersessian, N. (1992). Constructing and instructing: The role of 'abstraction techniques' in creating and learning physics. In R. Duschl & R. Hamilton (Eds.) *Philosophy of science, cognitive psychology and educational theory and practice* (pp. 48-68). New York: State University of New York Press.

Nersessian, N. (1993). In the theoretician's laboratory: Thought experimenting as mental modeling. In Hull, Forbes, & Okruhlick (eds.), pp. 291-301.

Nunez-Oviedo, M. C. (2003). *Teacher-student co-construction processes in biology: Strategies for developing mental models in large group discussions*. Unpublished Doctoral Dissertation, University of Massachusetts, Amherst, MA.

Nunez-Oviedo, M. C., Rea-Ramirez, M. A., & Clement, J. (in press).

Reiner, M. (1998). Thought experiments and collaborative learning in physics. *International Journal of Science Education, 20*, 1043-1058.

Reiner, M. & Burko, L. M. (2003). On the limitations of thought experiments in physics and the consequences for physics education. *Science & Education, 12*, 365-385.

Reiner, M. & Gilbert, J. (2000). Epistemological resources for thought experimentation in science learning. *International Journal of Science Education, 22*(5), 489-506.

Sorensen, R. A. (1992). *Thought experiments*. Oxford: Oxford University Press.

Stephens, L. & Clement, J. (2006). Using expert heuristics for the design of imagery-rich mental simulations for the science class. Proceedings of the NARST 2006 Annual Meeting, San Francisco, CA.

WGBH Videos. *Water wheels* P.O. Box 2284 S. Burlington, VT 05407-2284 PHN: (800) 255-9424. www.wgbh.org

Appendix A: Waterwheels transcript excerpt

1 S1: We did the variable of, like, the force of the water coming out of the tube, so we decided to lift up the bucket and so... so we figured the gravity pushed down the water so that it would come out faster out of the tubes and harder. Like

when we like when we did our table we just got a book on the table, regular, um we got 5 seconds with the cups. And then when we lifted up um, when we lifted it up a little bit we got 4 seconds. And, um, after we lifted it up higher, higher we got in lower until we got to 2 seconds. So that it was like about um where the ceiling was where we got about 2 seconds.

2 I: So your conclusion is...?

3 S1 Um, the higher the um the higher the bucket you lift it up, right, um, the more force, or the more force that comes out of the tube.

4 I: And the more force that comes out of the tube...?

5 S1: The um, faster, the um harder it hits the wheel. The harder it hits the wheel the faster we will turn, the faster we will turn the faster, the um, hi-, the faster the nails will come up.

6 I: So once again we are talking about a greater force, right? We are talking about a greater force and so the amount of H2O through the hose, [S1's] group is saying, increases the amount of force that water is hitting with, so what is another way of increasing the force that the water is hitting with?

7 S1: Oh, oh to increase the forces? Um, the height.

8 I: Okay, so [writing on the board] height... increases... force of ... water.

[S1} asked me, [S1} ask me if I had a longer hose, because he wasn't being very accurate, he didn't feel he was accurate when he was hitting the wheel?

9 S2: No... I saw that too because...

10 I: Do you think, do you think if you still had a longer hose, if it were traveling through the hose at that distance that you would have gotten the same results?

11 S3 Oh! No! OOOhh OOO OOO!

S1: [Probably not.]

12 S1: No, it wouldn't [S].

13 S2: I think because. I think the gravity would have had the same pull on it... even though it was out in the air. It has that gravity, that pull. And I think that even though if it is in the tube I think the pull is the same.

14 I: [S4]?

15 S4: I think the force would have been greater. It is sort of like a roller coaster when you go down the force is a lot greater when you go down than when you go up. So like the longer, if it's longer, right, it's faster... like it will go faster and the force will be greater.

16 I [S1]?

17 S1 I think the force, if the line is shorter, it would be more. Cause like say, say for example you're going down a slide, right, and if you, if you are going out a tall slide that is going off like, Lincoln School or something like that, it could be like real steep like this or something like that, you'll go slower, because you're, you are hitting the slide and the slide is slowing you down so it's like the friction. Right? But if you just drop off Lincoln School you go a lot quicker ... [laughing in background]. You go a lot quicker and you'll be bust up more than if you go out of the slide off of it.

18 S2 What do you mean by.... but like how...

19 S1 Cause look... cause look , listen..

20 I Hold on... hold on... go ahead [S1].

21 S1 I say the slide is going like this, right... yeah, say it's going- it could be going straight down, say the slide's all around you right? Like a tube slide, it's going straight down. When you're going down it, like your arms is hitting the sides of the tubes and stuff like that, and keep on bumping into it. Just like the water bumps into the sides of this tube right? So it will slow you down because of friction, right? and- but if you just drop, right? the only friction you'll be getting will be from the air and the air wouldn't cause as much as from the tube because the tube is a solid and the air's a gas.

22 I Will people agree that there is probably some friction within the tube?

23 S2 Yeah.

24 I OK. That's what's [S1's] saying. And friction, as we've studied before, will slow things down.

Appendix B: Gravity Transcript excerpt (coded for gestures)

173	S4: But I'm saying I think that almost all of gravity is done by the relation between mass and the *(xxx)*. There's an incredibly huge mass below us. All I've been saying from the beginning is that rotation has something to do with it too.			
174	T: Ok. Yeah.			
175	S4: I'm not saying that rotation is the only effect.			
176	T: Yeah. Ok. (S6)?			
177	S6: I was gonna ask how *(xxx)*.			
178	T: He's saying that the thing called gravity is a combination of two effects. Or at least two. Yeah?			
179	§<947851>S5: (S4), you said that ***[[holds right hand up, thumb and forefinger forming a circle]]*** rotation, it can ***[[brings hands together, apart, and back together, fingers pinched as if holding the ends of something]]*** also pull?	4. Motion depicts dynamics	7. Uses force term	
180	S4: I don't know. No, I didn't say that. Don't interpret what I said, 'cause it might have been....			
181	S5: Well that's literally what you said. I just don't think that it could *(xxx)*.			
182	S4: Ok say that it's throwing you. Then that still means that the top is still gonna be throwing you left and at the side(?). So your weight's gonna be different. *[[S4 is out of camera range during this utterance.]]*		7. Uses force terms	
183	S5: Yeah, but that has nothing to do with gravity.			
184	§<969472>S4: Why not? What if- the ***[[holding right forefinger vertically upward, enscribes a horizontal circle with his forearm, using his elbow as a pivot; turns his eyes away from S5 and appears to stare off into space]]*** Earth is trying to throw you around at the equator.	4. Motions depict dynamics	7. Uses force term	6a. Personal movement projection
185	S5: It's sort of like saying *(turning)* friction into a normal force and *(another force...)*.			
186	S4: How could that not have anything to do with it? If the Earth is trying to throw you off, in effect, at the equator, then it will kind of ***[[moves his right hand downward twice in a short, emphatic, vertical trajectory]]*** counteract the ***[[brings fists up in front of his face and pulls them sharply downward twice, as though pulling down on a bar]]*** pull of the Earth on you. And at the ***[[raises right hand above his head and points downward with both thumb and forefinger]]*** North Pole it wasn't trying to throw you off and the Earth has ***[[brings fists up slightly higher than before and pulls them sharply downward twice, as before, but with a slight bounce at the end]]*** more pull on you. Which means you'd weigh more. So it would change your weight.	3. Depictive motion (indicates location of North Pole as though it is in the air a few inches above his head). 4. Motions depict dynamics (3 motions)	7. Uses force term	6a. Personal movement projection
187	S5: I know, but I'm saying it has nothing to do with gravity. It changes your weight but it doesn't *(xxx gravity)*.			