Student Thinking about Static Equilibrium: Insights from Written Explanations to a Concept Question

JEFFREY L. NEWCOMER
Department of Engineering Technology
Western Washington University

PAUL S. STEIF
Department of Mechanical Engineering
Carnegie Mellon University

ABSTRACT

Students' thinking during an assessment task was monitored by analyzing their explanations of answers to a multiple-choice question about the equilibrium of bodies under given loads. Based on coding by independent raters, we found that nearly all students explained answers by appealing to force and/or moment summation; they differed principally on whether they applied both force and moment equilibrium and did so consistently. Correlations between answer selection and explanation as found from coding suggested that a reasonable guess as to a student's thinking can be made based on the answer choice alone. By the final exam, some students improved in their ability to consistently apply both principles or at least to apply them when needed. Many other students apparently learned to apply one of the principles consistently, although they came to ignore the other principle, even if they had acknowledged it at the start of the course.

Keywords: assessment, static equilibrium, student thinking

I. INTRODUCTION

One hallmark of the efforts to improve education and learning is a focus on the learner. This includes an uncovering of student thinking at a deeper, more detailed level, and how that thinking evolves over time and compares with experts. Many efforts to probe student thinking have been undertaken in science. Often, as reported in the Physics Education Research (PER) literature, for example, students are asked a question about a physical situation, and their reasoning and responses are probed (Bowden et al., 1992; Halloun and Hestenes, 1985; McDermott, 1984). Student response patterns are then used to infer underlying conceptual knowledge. Alternatively, one could explore the thought process of students as they explain their solutions to problems that are more typical of homework or exam problems (Streveler et al., 2006).

At the same time, there has been increased attention on assessment. Assessment is recognized as critical to the learning process (Black and William, 1998), allowing the student and instructor to focus attention on those areas where weakness is detected. Assessment also plays a role in assuring that educational goals are met. Assessment can focus on any skills or knowledge that is important to learning goals; assessment of conceptual understanding in particular has received a great deal of attention. With the Force Concept Inventory as a model (Hestenes, Wells, and Swackhamer, 1992), this has led to the development of concept inventories in engineering and other fields of science (Evans et al., 2003).

While there has been research dedicated to uncovering student thinking in a deep way, as well as research efforts aimed at developing efficient and meaningful assessment procedures, these activities should complement one another. That is, assessment procedures must be coherent: the assessment tasks should require students to draw upon the types of knowledge and skills that are central to the subject being learned. One framework for uniting these often parallel research activities is the assessment triangle (National Research Council, 2001). One apex of the triangle consists of the actual tasks carried out by persons being assessed; these are, for example, the questions to be answered or problems to be solved, which constitute the opportunities to observe the learner. The second apex is the model of cognition; that is, a model of what someone must think about in order to successfully complete the task, as well as the naïve or incomplete knowledge that non-experts typically bring to the task. Finally, the third apex is the means employed to interpret the observations of the assessment task. These three aspects should ideally be consistent with each other (Shavelson et al., 2002). That is, we should be aware of the thinking that students actually engage in during assessment activities, and their thinking should be consistent with learning objectives and instructional activities.

The study of student thinking in subjects similar to Statics, the domain targeted here, is an important component of PER. Two quite different ways of describing students' naïve knowledge have emerged. One view is that students enter courses with stable alternative conceptions (or misconceptions) of some physical phenomenon that reflect an internally consistent, if flawed, model; this model must be unseated before the student can learn scientifically correct thinking (for example: McCloskey, 1983; Gopnik and Wellman, 1994; Carey, 1999; Vosniadou, 2002). Another view is that students' incoming understanding is made up of bits of loosely connected knowledge referred to as phenomenological primitives (p-prims) (diSessa, 1993; Elby, 2000; Hammer, 1996a; Hammer, 1996b; Redish, 2004) or facets of thinking (Minstrell, 1992). These p-prims, or knowledge elements, are triggered by context; while they may individually be correct in some circumstances, they may be applied in the wrong circumstances or combined improperly to form incorrect conclusions. According to this view, students do not need to have misconceptions unseated; instead they must learn which elements are correct and in which contexts. These elements provide building blocks for reaching the desired mode of thinking. In this study, we do not intentionally seek data that sheds light on the competition between these views, nor could our observations of students in a limited context single out one position or the other. However, as will be seen, this study does reveal that students bring to Statics critical cognitive resources, though they exhibit "specific difficulties" (Loverude, Kautz, and Heron, 2002) and fail to deploy those resources properly.

The present study seeks: (i) to gain greater insight into the considerations students cite when addressing a challenging multiple-choice question frequently used for assessment that deals with the concept of static equilibrium, (ii) to determine how the considerations they cite relate to their answer choice for this question, and (iii) to monitor how the considerations they cite evolve with instruction. At issue in the equilibrium of bodies of finite size are the propensities of forces to cause rotational as well as translational motion. While student understanding of the relation between force and translational motion has been extensively studied, there have been relatively few attempts to study rotational effects. In Ortiz, Heron, and Shaffer (2005), students were probed regarding objects that rest on a pivot; the moments (torques, in the language common to physics) due to weights on the two sides must be equal. Students confused forces and moments: they mistakenly believed that forces on the two sides of the pivot, rather than their moments, had to be equal. Students also had trouble extending their ideas to objects with different orientations on the pivot, or believed that a tilted orientation was due to an unbalance in the forces (or moments). In Rimoldini and Singh (2005) students were probed on situations involving rotation and rolling. As in other work in PER, the authors detect a failure to appreciate that acceleration, not a constant velocity, accompanies an unbalanced force. The phenomena considered in Rimoldini and Singh (2005) are relatively complex, however, in that some of the forces are subtle ones, such as air drag; in addition, friction plays a role, and the complicated concept of moment of inertia is also at issue.

The method typically adopted in PER, such as the work in Ortiz, Heron, and Shaffer (2005) and Rimoldini and Singh (2005), is to focus student attention on a physical situation. Students are asked to predict and explain what will happen or to make qualitative statements regarding various parameters. These are generally conducted in the form of a recorded interview. These are referred to as "think-aloud" protocols, with students asked to explain their thinking, in contrast to merely a talk-aloud (used in psychological research), in which the subject is given a task to perform and is simply asked to talk aloud while engaged in the task (Ericcson and Simon, 1993). The talk-aloud is viewed as revealing what is in working memory and does not, in principle, divert the user from the normal way of performing the task. The performance of a task might be altered, however, if subjects are self-consciously seeking to explain themselves. The method used here is more in the style of the think-aloud, since we are asking explicitly for an explanation.

The question used in this study derives from a widely used and extensively analyzed assessment instrument, the Statics Concept Inventory (SCI) (Steif and Dantzler, 2005; Steif and Hansen, 2006; Steif and Hansen, 2007). Development of this instrument has been led by the second author. This inventory, which consists of 27 multiple-choice items, testing 9 concepts central to Statics, has been administered at over 25 institutions, often at the start of Statics, at the end of Statics or in follow-on courses. Among 1,164 students taking this test in fall 2005 either at the end of Statics or in follow-on courses, the percentage who correctly answered the question to be used in this study was 15 percent. This percentage is remarkably consistent from year to year and significantly less than the percentage correct on the test overall (49 percent) (Steif and Hansen, 2007). Other analyses of this test question, including correlation with other items in the same concept and discrimination based on item response theory, suggest it to be of high quality. In the following sections, we present the methodology used, including the question and the task posed to participants, the participants in the study, the context in which they completed the task, and how the results were analyzed. Results of the analysis are then presented, and some implications for instruction are drawn. Preliminary results from this study were reported in Newcomer and Steif (2006).

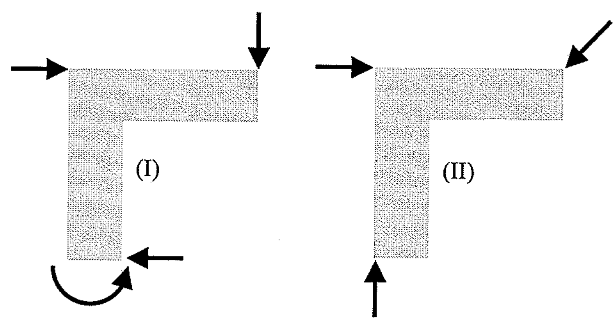
II. METHODOLOGY

A. Task

The multiple-choice question in Figure 1, an item from the SCI, was used in this study. Students were asked to choose an answer, and to explain in writing as fully as possible why the right answer is right and/or the wrong answers are wrong.

Equilibrium in Statics corresponds to zero net tendency to cause translational or rotational acceleration. Case (I) cannot be in equilibrium because there is nothing to balance the vertical force (which causes translational acceleration). Case (II) cannot be in equilibrium because the forces as they are drawn would cause a net moment, even

The forces and couple in the two cases shown act at the points indicated. All magnitudes are greater than zero, and the forces and couple act in the directions and senses shown.



Assuming the magnitudes of the forces and couple have the right values, could these bodies be in equilibrium?

(a) I could be in equilibrium; II could be in equilibrium
(b) I could never be in equilibrium; II could never be in equilibrium
(c) I could be in equilibrium; II could never be in equilibrium
(d) I could never be in equilibrium; II could be in equilibrium
(e) Cannot say without more information

Figure 1. Question from the SCI addressing equilibrium used in present study.

though force equilibrium can be satisfied if the forces have the right magnitudes. Because both the translation and rotation of the body are in question, the task is in some sense more complex than that studied in Ortiz, Heron, and Shaffer (2005) (where only rotation is at issue because of the pivot). However, the task is simpler than that in Rimoldini and Singh (2005) because the forces are given more explicitly (vector forces are drawn in the specific task here), and because dynamics and moments of inertia are not at issue.

Clearly, equilibrium is very fundamental to Statics; other than the fact that the item does require knowledge of the symbol of the couple and what it represents, no special or esoteric knowledge is required to answer this question. We recognize that students continue to struggle with the idea from physics that an unbalanced force causes acceleration, not constant velocity; likewise an unbalanced moment causes rotational acceleration, not constant rotational velocity. However, to simplify the language of this paper, we will henceforth use the shorthand of referring to forces as causing translation and rotation, with it understood that these are accelerations. We note that only four of 70 students in this study referred to accelerations at all in any of their explanations, and then only in regard to the linear acceleration in case I.

The approach pursued here to uncover student thinking departs slightly from what is typical in PER. In particular, rather than leaving students free to answer the question as they see fit, we have asked students why different answer choices are right or wrong. For many multiple-choice questions, such an approach would indeed run the risk of channeling student thinking and missing valuable information from students who think the answer lies outside the given options. However, in this situation with the question of whether the bodies are in equilibrium, the multiple-choice options given are exhaustive. There is no substantive difference (though, possibly a stylistic one) between the approach pursued here and a typical PER approach. The answer space that students are allowed to explore is sufficiently wide to give free range to their possible thought processes regarding equilibrium.

The methodology used here also differs from typical work in PER in that our data were obtained from written answers submitted electronically, not from interviews. Thorough probing of students in interviews would likely yield additional insights into their thinking. However, in support of the current approach, research on students in an introductory electricity and magnetism course (Scott, Stelzer, and Gladding, 2006) did find the results of multiple-choice exams, written explanations, and interviews to be strongly correlated. Moreover, aside from the pros and cons of different approaches to probing student thought, the approach pursued here allows one to ascribe greater meaning to responses to questions from concept inventories.

B. Participants

Data for this study came exclusively from students in a ten-week Statics course offered in the Winter 2006 quarter in the Engineering Technology (ET) Department at Western Washington University (WWU). Students performed the task twice: in the second week of the course prior to any instruction on equilibrium and again on the final exam (where the two cases were swapped and flipped horizontally, leaving the question conceptually identical). Thus, the data consist of one set of answer choices, each with its corresponding explanation, from early in the quarter, and a similar set from the final exam. As described below, not all students in the class participated in this exercise in both instances.

The Statics course at WWU, which has a pre-requisite of one quarter of physics (mechanics), is organized around five topics: free body diagrams, equilibrium, equivalence, separation of rigid bodies, and friction, without differentiation between two vs. three dimensional cases, concurrent vs. non-concurrent force systems, and single bodies vs. frames and trusses (Newcomer, 2006). These situations are all addressed in the course, but not in the order of traditional textbooks. Otherwise the course is a standard lecture-based course with homework, midterms, a design project, a final exam, and a limited number of think-pair-share exercises. Students are also asked to complete weekly warm up exercises for the first eight weeks of the quarter. These exercise touch on a variety of topics, such as free body diagrams (FBDs), equilibrium, equivalence, separation of bodies at connections, and friction (Patterson and Novak, 1999–2006).

For each warm up exercise, students are asked to answer a given multiple-choice question from the SCI and to explain the reason for their selection. In this case, the purpose of the warm up exercises is to give the instructor an idea of the state of student understanding of the topic in order to help guide instruction (Heron, 2004a; Heron, 2004b). Each week students are given a period of slightly more than 24 hours during which they must access the question and submit their answers and explanations, all through an electronic course management system. Those who answer the question and give an explanation receive full credit for the assignment regardless of whether their answers are correct or not. Since students respond to this assignment on-line, do not receive additional credit for giving the correct answer, and have not formed homework working groups near the start of the quarter (when the particular warm up exercise studied here was posed), we expect instances of student collaboration on the warm up exercise in question here to be rare, and not to significantly affect the findings of the study. During lecture the course instructor uses the answers and explanations from the weekly warm up question to address specific difficulties that proved to be common. Students are then asked to answer similar questions and provide explanations for their selection on the final exam.

C. Analysis

In addition to tabulating and analyzing answer choices, we sought to categorize or code responses based on a reading of an entire explanation and on the answer chosen. A coding scheme was devised based on preliminary data that were obtained during the Winter 2005 academic term from a group of 39 students at WWU. An iterative process was followed in approaching the preliminary data: an initial set of categories (codings) was proposed, both authors applied the coding scheme to the data set, discrepancies were discussed, and the set of categories was revised. The process continued until near convergence was reached. After the scheme was completed based on the preliminary data, the same scheme was applied essentially unaltered by the authors independently to responses from participants in the present study (70 total students in the Winter 2006 Statics class at WWU). For greater clarity, the major categories presented in Newcomer and Steif (2006) have been renamed.

The coding scheme is:

- Force Equilibrium
  1. Never assessed
  2. Assessed, but not when needed
  3. Assessed just when needed
  4. Always assessed

- Moment Equilibrium
  1. Never assessed
  2. Assessed, but not when needed
  3. Assessed, just when needed
  4. Always assessed
- Acknowledged that forces cause rotation (Y/N)

The coding "Never assessed" means that the associated equilibrium principle was not included in the explanation. The coding "Assessed, but not when needed" means that the associated equilibrium principle was not applied in the case when it was critical to the correct answer, although the principle was mentioned in reference to the other case. The coding "Assessed, just when needed" means that the principle was invoked in the one case when it was needed to arrive at the correct answer, but not in both cases. Finally, the coding "Always assessed" means that the principle was invoked in both cases, even though it was necessary to the correct answer in only one case. Thus, we would treat the knowledge described by coding 2 as superior to that described by coding 1, and that described by coding 3 as superior to that described by coding 2. Given the task and that each case can be judged as not in equilibrium by invoking a single principle, we argue that coding 3, invoking the principle only as necessary to justify the answer choice, does not necessarily imply inferior knowledge to that implied by coding 4, invoking of the principle in both cases.

The following are examples of typical student explanations, along with the associated codings:

Answer (a)
Explanation:
1: if you look purely at the forces it is trying to rotate the object clockwise and the couple is trying to rotate the part counter clockwise, so it could be in equilibrium.
2: could be in equilibrium because there are X and Y component forces that could counter the angled force.
Coding: Force equilibrium – 2, Moment equilibrium – 2, Forces cause rotation – Y

Answer (b)
Explanation:
I could not be in equilibrium because the linear forces do not cancel out. The horizontal forces do, but not the vertical.
II could not be in equilibrium because, although the linear forces could cancel out, the moment caused by the vertical components act in the same direction and cannot cancel one another out.
Coding: Force equilibrium – 4, Moment equilibrium – 3, Forces cause rotation – Y

Answer (c)
Explanation:
1 can be in equilibrium, because all of the forces, excluding the moment, will cause a clockwise rotation, which can be counteracted by the moment.
2 cannot be in equilibrium, because the y-component of the diagonal force and the lower left hand force will always cause a clockwise rotation.
Coding: Force equilibrium – 1, Moment equilibrium – 4, Forces cause rotation – Y

Answer (d)
Explanation:
I don't think system 1 could be in equilibrium, because there is a negative y force acting that doesn't have anything pushing back on it.
I think that 2 could be in equilibrium, since one of the forces opposes the other two at an angle that it could equal out the x and y forces.
Coding: Force equilibrium – 4, Moment equilibrium – 1, Forces cause rotation – N

The authors independently coded the student explanations that went with the answer selections. The results reported below use the average of the two codings. The quality of the codings, or their validity, should be judged on the extent to which the multiple individuals carrying out the coding task (the two authors) agree. While the fraction of instances where raters agree could be a reasonable measure, the so-called inter-rater reliability is more commonly assessed in social and medical science research using the statistic κ, rather than mere agreement. The statistic κ discounts the amount of agreement that might be due purely to chance (Chest X-Ray, 2000; Kundel and Polansky, 2003). Agreement by chance can occur when reviewers know that a certain outcome, say a null result, is more likely, and then assign unclear cases that value. Agreement in these cases is expected. The statistic κ is then the ratio of (overall agreement minus expected agreement) to (total cases minus expected agreement). As a result, the value of κ will be no greater than the fraction of instances in which the raters agree, but it is a more accurate measure of inter-rater reliability than mere agreement.

III. Results

The distributions of student answer selections for the warm up and final exam are shown in Figure 2. Two students who completed the warm up did not complete the final exam, so a comparison between the warm up and the final exam answers for individual students is possible for only 56 of the responses.

As Figure 2 shows, some students in this class have progressed: far more have chosen answer (b) at the final exam as compared to the warm up, recognizing cases I and II both to be out of equilibrium.

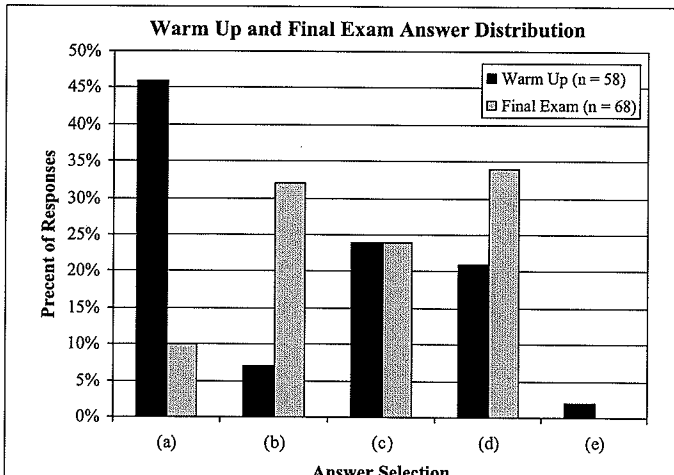


Figure 2. Student answer selection distribution on the warm up and final exam. Answer (b) is correct.

Warm Up Answer	Final Exam Answer a	b	c	d
a (n = 25, 46%)	3 (12%)	8 (32%)	6 (24%)	8 (32%)
b (n = 4, 7%,)	1 (25%)	3 (75%)	0 (0%)	0 (0%)
c (n = 14, 25%)	1 (7%)	4 (29%)	3 (21%)	6 (43%)
d (n = 12, 22%,)	1 (8%)	4 (33%)	3 (25%)	4 (33%)

Table 1. Distributions of student answer selections on the final exam for different selections on the warm up.

The frequency of wrong answers (a) decreased sharply, although wrong answers (c) and (d) remain popular.

By retaining the association between each student and his or her answer selection, the change over the term could be tracked. The relation between answer choices on the warm up and the final exam for the 55 students who took both and did not answer (e) on the warm up are shown in Table 1.

In Table 1, one observes that the distribution of students giving the respective answers (a), (b), (c), and (d) on the final appear to be roughly similar, regardless of the answer given by students at the start of the course (aside from the four who answered correctly on the warm-up). We sought to determine if the distributions are statistically different using the chi-squared double distribution test (Deacon, 2007), which can be used to determine if two distributions are statistically different from each other in circumstances where there is no expected distribution for either one. According to this test, the distributions on the final exam are extremely similar ($p = 0.73$) for those with warm up answers (a) and (c), very similar ($p = 0.62$) for those with warm up answers (a) and (d), and similar ($p = 0.51$) for those with warm up answers (c) and (d). Still, given the small numbers, we cannot conclude with certainty that the distributions are statistically the same. That uncertainty aside, though, the answer choice on the warm up is surely not a predictor of the answer choice on the final exam.

The distribution of codings for both the warm up and final exam, averaged for the two raters, are given in Tables 2 and 3 for force equilibrium and moment equilibrium, respectively. (Because of the averaging process and disagreement between raters, results here and below sometimes tally to one-half student.) Also appearing in these tables are the inter-rater reliability coefficients, κ. According to Chest X-Ray (2000), the range $0.41 < \kappa < 0.60$ indicates moderate agreement, $0.61 < \kappa < 0.80$ indicates significant agreement, and $0.81 < \kappa < 1.00$ indicates near perfect agreement. In general, agreement between raters is in the range from significant to near perfect.

It can be seen that some explanations were found to be uncodeable, meaning that one or both raters could not assign a coding score to the student's response; there were fewer uncodeable explanations in the final exam compared with the warm up. Note that the coding scheme captures whether students *attempt* to apply a condition of equilibrium, force or moment, not whether the condition was applied correctly. For that reason, we also tracked instances in which students applied a principle, but did so incorrectly, as judged by reaching a wrong conclusion; we catalog this as a misapplication. The misapplication of force equilibrium declined from 5 percent at the warm up to 2 percent at the final exam, and the misapplication of moment equilibrium declined from 16 percent at the warm up to 5 percent at the final exam.

Force Equilibrium	Warm Up n = 58, 4 correct		Final Exam n = 68, 22 correct	
Coding	Number of Responses	κ	Number of Responses	κ
1	17½ (30%)	0.877	11 (16%)	0.783
2	22 (38%)	0.926	8 (12%)	0.575
3	1½ (3%)	0.482	12 (18%)	0.732
4	10½ (18%)	0.888	34½ (51%)	0.911
Uncodeable	6½ (11%)	--	2½ (4%)	--

Table 2. Frequencies of codings and inter-rater reliabilities for force equilibrium.

Moment Equilibrium	Warm Up n = 58, 4 correct		Final Exam n = 68, 22 correct	
Coding	Number of Responses	κ	Number of Responses	κ
1	7½ (13%)	0.617	14½ (21%)	0.956
2	20½ (35%)	0.811	12 (18%)	0.898
3	1½ (3%)	-0.023	16 (24%)	0.874
4	25 (43%)	0.859	22½ (33%)	0.966
Uncodeable	3½ (6%)	--	3 (4%)	--

Table 3. Frequencies of codings and inter-rater reliabilities for moment equilibrium.

One critical issue is how explanations are related to answer choice. Table 4 shows the distributions of codings given for each answer choice in the final exam. In some cases the sums are less than the total number of students, because some statements were uncodeable. We sought to determine if the codings are correlated with answer selection, for example a coding of 4 on force equilibrium and a coding of 1 on moment equilibrium with answer (d), using an ANOVA. The different answers were treated as factor levels and the codings as the values of the responses; individual responses (codings) were then grouped by level (answer). The association of codings with answers are statistically significant ($p < 0.001$) for both force and moment codings. A similar table could be constructed for the warm up, and, although the pattern is less striking due in part to the higher rate of errors in applying force or moment equilibrium, the associations of codings with answer selection are also statistically significant ($p < 0.001$).

For brevity in the discussion, we will define a coding of F3 to be a coding of 3 on force equilibrium, a coding of M2 to be a coding of 2 on moment equilibrium, and so forth. The following insights regarding students' knowledge state at the end of Statics, as signaled by answer choice on the final, can be obtained from the distributions of codings for different answers, shown in Table 4:

> *Answer (a):* There is a strong presence of both codings F2 and M2, indicating that both principles are invoked, but not consistently and not when necessary. Thus, this answer choice indicates a modest level of understanding: one may consider force equilibrium in some cases and moment

Answer	Force Coding 1	2	3	4	Moment Coding 1	2	3	4
a (n =7)	½ (7%)	5½ (79%)	0 (0%)	0 (0%)	½ (7%)	5½ (79%)	0 (0%)	½ (7%)
b (n = 22)	½ (2%)	½ (2%)	11 (50%)	10 (46%)	1 (5%)	0 (0%)	11 (50%)	9 (41%)
c (n = 16)	10 (62%)	2 (12%)	0 (0%)	2½ (16%)	0 (0%)	0 (0%)	2 (12%)	13 (81%)
d (n = 23)	0 (0%)	0 (0%)	1 (4%)	22 (96%)	13 (57%)	6½ (28%)	3 (13%)	0 (0%)

Table 4. Comparison of answer selection and coding on final exam.

Answer	Force	Moment
(a) I could be in equilibrium; II could be in equilibrium	Ignored selectively	
(b) I could never be in equilibrium; II could never be in equilibrium	Assessed always or when needed	
(c) I could be in equilibrium; II could never be in equilibrium	Selectively/ Never used	Always assessed
(d) I could never be in equilibrium; II could be in equilibrium	Always assessed	Selectively/ Never used

Table 5. Descriptive summary of consideration given to force and moment balance for different answer selections.

equilibrium in other cases, but not both and not consistently. Such a student might be termed a "jack of all principles, but a master of none." One might speculate that such students are provoked into thinking about moment equilibrium only when seeing the couple and about force equilibrium only when seeing only forces, although this speculation could not be verified from these data.

Answer (b): There is a strong presence of codings F3 or F4 and codings M3 or M4. Both principles are invoked and applied at least when necessary to arrive at a correct conclusion, if not all the time. Since one only need disprove one principle to show a case is not in equilibrium, a student might, for example, assert that case I violates force equilibrium, and then move on to case II without discussing moment equilibrium for case I. Given the task, it would be hard to claim that students who received a coding of F3 or M3 had any weaker understanding than those who received a coding of F4 or M4. In fact, the cases of coding 1 or 2 for this correct answer choice corresponded to the only cases in which the explanation suggested clearly that the students intended an answer other than (b) as their choice.

Answers (c) and (d): Nearly all of these students have demonstrated a mastery of consistently applying either force equilibrium (d) or moment equilibrium (c), but not both. Of those who chose (d), mostly coded by F4 and some by F3, 57 percent were coded M1, and only 28 percent were coded M2. Of those who chose (c), mostly coded by M4 and some by M3, 62 percent were coded F1, and only 12 percent were coded F2. Thus, in both cases, the majority did not even mention the insufficiently attended principle at all. This majority of the students choosing (c) or (d) might be termed "narrow specialists." By consistently applying at least one principle, these students are advanced relative to those who select (a); however, they are weaker than those who chose (a), in that they completely ignored the other concept. Most of those who chose (a) used both concepts, albeit insufficiently.

We summarize these associations between answer selection and explanation in Table 5.

Aside from the answer choices, one can use the distribution of force and moment codings to draw additional conclusions regarding conceptual progress of the students as a whole in this course. Figures 3 and 4 show the force and moment coding distributions, respectively, for those answers that were codeable. These figures combine students with codings 3 and 4 into one group, consistent with the argument above that these distinct codings do not necessarily imply different levels of knowledge, given the task. Perhaps surprisingly, relatively few students (21 percent) apply force equilibrium as necessary at the start, whereas nearly half (46 percent) apply moment equilibrium as necessary. Instruction clearly produces more progress with respect to force equilibrium than moment equilibrium: both of the codings 1 and 2 decrease for force, and the codings 3 and 4 increase dramatically. By contrast, more students never mention moment equilibrium at the final as compared with the warm-up, and the increase in codings 3 and 4 for moment equilibrium is relatively modest (46 percent to 57 percent). The

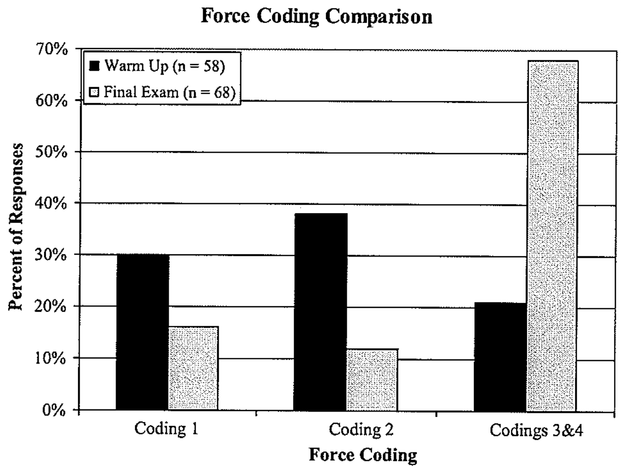


Figure 3. Comparison of coding at warm up and final exam for force.

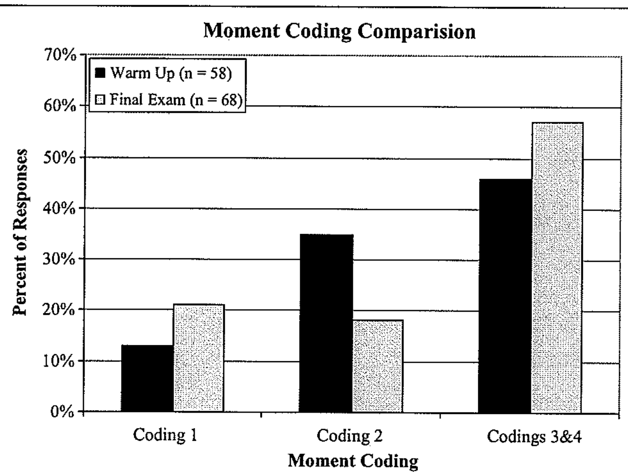


Figure 4. Comparison of coding at warm up and final exam for moment.

conclusion regarding the greater proficiency in the application of force equilibrium at the time of the final could also have been arrived at based on answer choices alone: more students choose (d), which ignores moment equilibrium, than choose (c), which ignores force equilibrium.

Tables 6 and 7 show the coding progression of individual students from warm up to final exam for force and moment equilibrium, respectively. The data include the 55 students who took both the warm up and the final exam, did not choose (e) on the warm up and had warm up answers that were codeable. Note that, due to uncodeable responses on the final exam, the rows do not all add up to 100 percent. While many students improved from the warm up to the final exam, a significant fraction of students who were thorough in either force or moment equilibrium (coding 4) at the warm up, apparently regressed by the final, and failed to mention the condition at all or applied it insufficiently. Included in this group are nine students who flip-flopped entirely from the warm up to the final exam. These students switched from codings of F1 or F2 and M4 on the warm up to codings of F4 and M1 or M2 on the final exam or vice versa.

Along with judging the application of force and moment equilibrium, we also captured whether students gave evidence of recognizing that forces play a role in rotation at all. While the majority explicitly acknowledged such a role, a distinct minority gave explanations that made no reference to this role (no student explicitly asserted that forces do *not* cause rotation). The percentages which acknowledged this role are tabulated in Figure 5 for each of the answer choices. Overall, 72 percent of students (n = 58, κ = 0.913) acknowledged that forces cause rotation on the warm up, and 68 percent of students (n = 68, κ = 0.966) did so on the final exam.

Warm Up	Final Exam Force Coding		
Force Coding	1	2	3 & 4
1 (n = 16½, 33%)	3 (18%)	2½ (15%)	10½ (64%)
2 (n = 21, 43%)	3 (14%)	1½ (7%)	16½ (77%)
3 (n = 1½, 3%)	0 (0%)	0 (0%)	1½ (100%)
4 (n = 10½, 21%)	2 (19%)	2 (19%)	5½ (52%)

Table 6. Force coding distributions on the final exam based on force coding on the warm up.

Warm Up	Final Exam Moment Coding		
Moment Coding	1	2	3 & 4
1 (n = 7½, 14%)	1½ (20%)	1 (13%)	4 (53%)
2 (n = 20½, 39%)	8 (39%)	½ (2%)	12 (59%)
3 (n = 1½, 3%)	0 (0%)	0 (0%)	1 (67%)
4 (n = 23, 44%)	2 (9%)	6 (26%)	14½ (63%)

Table 7. Moment coding distributions on the final exam based on moment coding on the warm up.

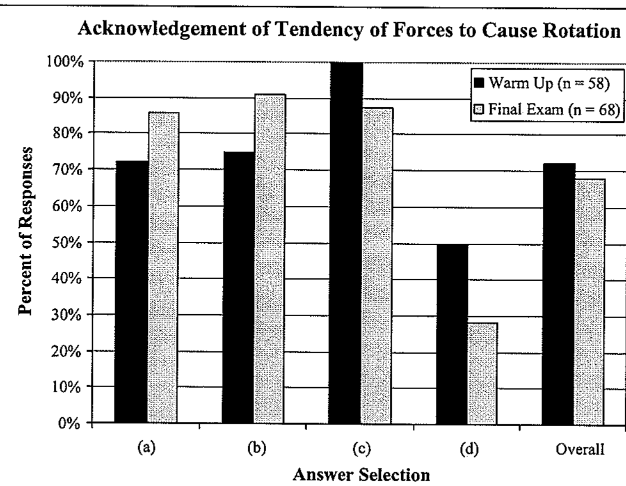


Figure 5. Percentage of students that explicitly acknowledged the tendency for forces to cause rotation.

Not surprisingly, on both the warm up and the final exam, students who received a coding of M1 did not acknowledge the role of forces in rotation, while students who received a coding of M4 always acknowledged the role of forces in rotation. Also, not surprisingly, the percentage that acknowledges the role of forces in rotation is least for those who answered (d), which can indeed be arrived at without consideration of moments.

IV. General Discussion

Virtually all students at both the start of Statics and at the final exam use language that invokes the relevance of combining forces and/or combining of moments (or rotational tendencies). Furthermore, when such a principle is applied, the student usually arrives at a correct conclusion, and the number of incorrect applications of a principle does decrease from the start to the finish. However, many students, even at the final exam, do not apply *both* force and moment equilibrium *consistently*, or at least when they are needed. While a larger sample size would perhaps give further assurance, the results of this study strongly suggest these conclusions.

Thus, in the language of Hammer (1996a; 1996b), which reflects a philosophy in common with diSessa's (1993) proposition regarding the p-prims possessed by novices, students come to Statics with substantial cognitive "resources", which can be activated as needed to solve problems. It is true, as Ortiz, Heron, and Shaffer (2005) documented that students facing the beam with weights on the two sides of a pivot often have trouble grasping that moment, rather than force, balance is critical. Consistent with her finding, we also found students, even at the end of Statics, less likely to invoke moment equilibrium than force equilibrium. Furthermore, based on our coding for the recognition that forces had any role to play in rotations, there is no progress at all in this respect, with a substantial minority still not volunteering the presence of any such a role.

With respect to students entering Statics, the resources they arrive with are, by and large, the *most* that one could expect from instruction in physics. For the beam on a pivot, which is the primary exposure students have, if any, to moments (torques) in physics instruction, the equilibrium of forces is not even at issue: the pivot supplies any needed vertical force. Thus, students rarely have any exposure to the need to invoke force and moment equilibrium simultaneously. Statics, by contrast, demands the new skill of simultaneously juggling these two ideas. Judging by the percentage of students in this class (32 percent) who can do so, and the number who respond correctly to this item on the SCI nationwide (15 percent), Statics courses are by and large not successful in developing this skill. Without further study, we cannot answer the important question as to what is difficult to grasp about forces simultaneously producing translation and rotation that impedes the acquisition of this skill, even though the individual cognitive resources are present.

Do students start Statics with in-grained modes of thinking that prevent them from progressing? With the depth to which this study can probe, it would seem not. Since the distributions of answers on the final exam are at least approximately the same for students who answered differently at the start of the course, it does not appear difficult to change students' minds. While some students have progressed, others would appear to have regressed somewhat: they might have invoked a principle at the warm up which they failed to invoke again at the final exam. Such regression may be genuine, or it may imply that initial knowledge was inert—they can recite words to the effect of summing forces and moments, without fully grasping the implications.

V. Summary and Implications for Instruction

By analyzing the written explanations offered by students for their answers to a multiple-choice question from the Statics Concept Inventory (SCI), we have investigated the considerations that students bring to bear in approaching equilibrium. As judged by the sizable numbers of students who have taken the SCI online over a number of years, the particular question used for this study is known to be very difficult for students. We have found in this study that both at the start of a Statics course, and at the final exam, students by and large invoke the expected considerations of force and/or moment balance. This is not surprising, given the previous experience in physics. In fact, when students do invoke such a balance principle, they tend to apply it correctly. So many students answer this question incorrectly, we found, because they either fail to apply both force and moment equilibrium, or fail to do so consistently.

Based on coding of the explanations, student invocation of each principle (force and moment) was described as follows: (1) never, (2) insufficiently, (3) appropriately, just as needed, or (4) in all cases. The answer choice was found to be strongly correlated to the coding, particularly at the final exam, but also at the warm up. That is, nearly all students answering correctly that neither case was in equilibrium invoked both force and moment at a level of (3) or (4). Students who answered that only one case was not in equilibrium indeed invoked force or moment at the level of (3) or (4); specifically force was invoked when case I was declared to be out of equilibrium, and moment was invoked for case II. Students who thought both bodies were in equilibrium tended to have codings of (1) or (2) for invoking both force and moment equilibrium. Should these findings hold for larger samples of students across a range of institutions, then one could make a reasonable guess as to the thinking of a student from the answer choice on this item of the SCI.

We found a mixed pattern of progress with instruction in this group of participants. At the start of the course, many students alluded to both force and moment equilibrium, but did not do so consistently and so mistakenly identified both cases as in equilibrium. By the end of the course, many students became narrow specialists: they consistently invoke one principle, thereby identifying one case as out of equilibrium. But, they would disregard the other principle, identifying the other case as in equilibrium. There was a sizable number of students, nearly all of whom answered correctly, who managed to invoke both principles, as needed for the correct conclusion or consistently across cases.

What instruction could enable significantly more than one-third of students to answer this question correctly? It is not obvious that instruction needs to be tailored to the initial conceptual state of the learner, since the distribution of answers at the final was roughly independent of the answer at the start of the semester. But, instruction does have to acknowledge the propensity of so many students either to disregard one of the two conditions entirely or to incompletely apply at least one condition. In seeking more effective instruction, it should also be recalled that typical Statics problems require the creation of a free body diagram (FBD) and the imposition of equilibrium conditions on that FBD. There is the tacit assumption in typical problems that equilibrium is possible. Thus, the style of SCI question studied here—asking whether equilibrium is possible—is certainly different from the typical experience of students. On the other hand, for the process of design, one can imagine the benefit in being able to infer whether a contemplated arrangement could possibly be in equilibrium. Instruction that promotes the contemplation of the possibility of equilibrium, rather than only the quantitative implications of its presence, may both improve performance on this assessment task and may have broader benefits in preparing students for using mechanics in engineering practice.

We finally comment on the broader usefulness of the methodology used here to collect data. While they have been valuable for research purposes, should instructors more generally use warm up exercises to gain insight into student thinking on a topic before it is covered in class? For many subjects, student thinking is relatively unexplored and such warm ups can offer valuable guidance to instructors (who must devise their own questions). On the other hand, for subjects in which learning has been more thoroughly studied, the typical thinking of students, if broadly representative, can be established once and for all based on a few studies. An instructor could have students only select answers, particularly if explanations are known to correlate with answer choice (as found here). However, without the requirement to supply a reason, students might give far less thought to their answers. In addition, by asking students to explain, we may enhance their interest if and when the instructor later discusses the reasoning in class. Further, instructors probably need to give at least the impression of using the warm up responses in order for them to be more than a game to the students. In practice, of course, since the use of responses for discussion is usually done anonymously, the instructor could have typical responses in hand for discussion purposes, rather than undertaking the time-consuming task of reading and sorting all of the responses from a large class.

ACKNOWLEDGMENTS

Support by the National Science Foundation under grant REC-0440295 and by the Department of Mechanical Engineering at Carnegie Mellon University and the Department of Engineering Technology at Western Washington University is gratefully acknowledged.

REFERENCES

Black, P., and D. William. 1998. Assessment and classroom learning. *Assessment in Education* 5 (1): 7–73.

Bowden, J.J., G. Dall'Alba, E. Martin, D. Laurillard, F. Marton, G. Masters, P. Ramsden, A. Stephanou, and E. Walsh. 1992. Displacement, velocity and frames of reference: phenomenographic studies of students' understanding and some implications for teaching and assessment. *American Journal of Physics* 60 (3): 262–69.

Carey, S. 1999. Sources of conceptual change. In *Conceptual development: Piaget's legacy*, ed. E. Scholnick, K. Nelson, S. Gelman, and P. Miller, 293–236. Mahwah, NJ: Lawrence Erlbaum Associates.

Chest X-Ray. October 2000. Agreement observer. http://www.chestx-ray.com/Statistics/kappa.html.

Deacon, J. 2007. Chi-squared test for categories of data. *The Really Easy Statistics Site.* http://helios.bto.ed.ac.uk/bto/statistics/tress1.html#THE%20REALLY%20EASY%20STATISTICS%20SITE.

diSessa, A. 1993. Towards an epistemology of physics. *Cognition and Instruction* 10 (2 & 3): 105–225.

Elby, A. 2000. What students' learning of representations tells us about constructivism. *Journal of Mathematical Behavior* 19 (4): 481–502.

Ericcson, K.A., and H.A. Simon. 1993. *Protocol analysis*, (Rev.). Cambridge, MA: MIT Press.

Evans, D.L., D. Gray, S. Krause, J. Martin, C. Midkiff, B.M. Notaros, M. Pavelich, D. Rancour, T. Reed-Rhoads, P. Steif, R. Streveler, and K. Wage. 2003. Progress on concept inventory assessment tools. In *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference.* Boulder, Co.

Gopnik, A., and H.M. Wellman. 1994. The theory theory. *Mapping the mind: Domain specificity in cognition and culture*, ed. L.A. Hirschfeld and S.A. Gelman. New York: Cambridge University Press, 257–93.

Halloun, I.A., and D. Hestenes. 1985. The initial knowledge state of college physics students. *American Journal of Physics* 53: 1043–55.

Hammer, D. 1996a. Misconceptions or p-prims: how may alternative perspectives of cognitive structure influence instructional perceptions and intentions. *The Journal of the Learning Sciences* 5 (2): 97–127.

Hammer, D. 1996b. More than misconceptions: multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics* 64 (10): 1316–25.

Heron, P.R.L. 2004a. Empirical investigations of learning and teaching, part i: Examining and interpreting student thinking. In *Proceedings of the International School of Physics "Enrico Fermi" Course CLVI*, Varenna, Italy, ed. E. F. Redish and M. Vicentini, 341–350. Amsterdam: IOS Press.

Heron, P.R.L. 2004b. Empirical investigations of learning and teaching, part ii: Developing research-based instructional materials. In *Proceedings of the International School of Physics "Enrico Fermi" Course CLVI*, Varenna, Italy, ed. E.F. Redish and M. Vicentini, 351–365. Amsterdam: IOS Press.

Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force concept inventory. *The Physics Teacher* 30 (3): 141–51.

Kundel, H.L., and M. Polansky. 2003. Measurement of observer agreement. *Radiology* 228 (2): 303–08.

Loverude, M.E., C.H. Kautz, and P.R.L. Heron. 2002. Student understanding of the first law of thermodynamics: relating work to the adiabatic compression of an ideal gas. *American Journal of Physics* 70 (2): 137–48.

McCloskey, M. 1983. Naive theories of motion. *Mental Models*, ed. D. Gentner, and Stevens, 299–324. Mahwah, NJ: Lawrence Erlbaum Associates.

McDermott, L.C. 1984. Research on conceptual understanding in mechanics. *Physics Today* 37 (7): 24–32.

Minstrell, J. 1992. Facets of students' knowledge and relevant instruction. In *Proceedings of an International Workshop—Research in Physics Learning: Theoretical Issues and Empirical Studies*, ed. R. Duit, F. Goldberg, and H. Niedderer, 110–128. Kiel, Germany: The Institute for Science Education (IPN).

National Research Council. 2001. *Knowing what students know: The science and design of educational assessment*, ed. J.W. Pellegrino, N. Chudowsky, and R. Glaser. Washington, DC: National Academies Press.

Newcomer, J.L. 2006. Many problems, one solution method: Teaching statics without 'special cases'. In *Proceedings of the 36th Annual ASEE/IEE Frontiers in Education Conference*, S2D-7-12. San Diego, CA.

Newcomer, J.L., and P.S. Steif. 2006. Student explanation of answers as a window into prior misconceptions. In *Proceedings of the 36th Annual ASEE/IEE Frontiers in Education Conference*, S4H-6-11. San Diego, CA.

Ortiz, L.G., P.R.L. Heron, and P.S. Shaffer. 2005. Student understanding of static equilibrium: predicting and accounting for balancing. *American Journal of Physics* 73 (6): 545–53.

Patterson, E., and G. Novak. 1999-2006. *Just-In-Time Teaching*, http://webphysics.iupui.edu/jitt/jitt.html.

Redish, E.F. 2004. A theoretical framework for physics education research: modeling student thinking. In *Proceedings of the International School of Physics "Enrico Fermi" Course CLVI. Varenna*, Italy, ed. E.F. Redish and M. Vicentini, 1–50. Amsterdam: IOS Press.

Rimoldini, L.G., and C. Singh. 2005. Student understanding of rotational and rolling motion concepts. *Physical Review Special Topics—Physics Education Research* 1. 010102: 010102-1-9.

Scott, M., T. Stelzer, and G. Gladding. 2006. Evaluating multiple-choice exams in a large introductory physics course. *Physical Review Special Topics—Physics Education Research 2*. 020102: 020102-1-14.

Shavelson, R.J., M. Li, M.A. Ruiz-Primo, and C.C. Ayala. 2002. Evaluating new approaches to assessing learning. Invited paper keynote speech at *Joint Northumbria/EARLI Assessment Conference*, University of Northumbria at Newcastle.

Steif, P.S., and J.A. Dantzler. 2005. A statics concept inventory: development and psychometric analysis. *Journal of Engineering Education* 94 (4): 363–71.

Steif, P.S., and M.A. Hansen. 2006. Comparisons between performances in a statics concept inventory and course examinations. *International Journal of Engineering Education* 22: 1070–76.

Steif, P.S. and M.A. Hansen. 2007. New practices for administering and analyzing the results of concept inventories. *Journal of Engineering Education* 96 (3): 205–12.

Streveler, R.A., M.R. Geist, R.F. Ammerman, C.S. Sulzbach, R.L. Miller, B.M. Olds, and M.A. Nelson. 2006. The development of a professional knowledge base: The persistence of substance-based schemas in engineering students. *Annual Meeting of the American Educational Research Association*. San Francisco, CA.

Vosniadou, S. 2002. On the nature of naïve physics. In *Reconsidering conceptual change: Issues in theory and practice*, ed. M. Limón and L. Mason, 61–76. Dordrecht, Netherlands: Kluwer Academic Publishers.

Authors' Biographies

Jeffrey L. Newcomer is a professor of Manufacturing Engineering Technology at Western Washington University. He received B.S. (1988) and M.Eng. (1989) degrees in Aeronautical Engineering, a M.S. in Science and Technology Studies (1993), and a Ph.D. in Mechanical Engineering (1994) from Rensselaer Polytechnic Institute. He is engaged in research to improve instruction and assessment in engineering, with an emphasis on engineering fundamentals such as mechanics.

Address: Engineering Technology Department, Western Washington University, Bellingham, WA 98225-9086; telephone: (+1) 360.650.7239; fax: (+1) 360.650.4847; e-mail: jeff.newcomer@wwu.edu.

Paul S. Steif is a professor of Mechanical Engineering at Carnegie Mellon University. He received a Sc.B. in engineering from Brown University (1979) and M.S. (1980) and Ph.D. (1982) degrees from Harvard University in applied mechanics. He has been active as a teacher and researcher in the field of engineering mechanics. In particular, Dr. Steif develops and implements new approaches and technologies to measure student understanding of engineering and to improve instruction.

Address: Department of Mechanical Engineering, Carnegie Mellon University, Schenley Park, Pittsburgh, PA, 15213; telephone: (+1) 412.268.3507; fax: (+1) 412.268.3348; e-mail: steif@cmu.edu.