

COMPARISONS BETWEEN PERFORMANCES IN A STATICS CONCEPT INVENTORY AND COURSE EXAMINATIONS

PAUL S. STEIF

Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: steif@andrew.cmu.edu

MARY HANSEN

Department of Secondary Education and Graduate Studies, Robert Morris University, Pittsburgh, PA 15213, USA. Email: Hansen@rmu.edu

Multiple analyses of results are presented from the Statics Concept Inventory, a multiple choice test assessing conceptual knowledge in engineering Statics. Results are based on the administration of this test to 1331 students in ten classes at seven US universities during the 2004-2005 academic year. Evidence confirming the validity and reliability of the test is offered. Detailed comparisons are made between inventory scores and performance on course examinations, including evaluations of how inventory sub-scores on specific concepts correlate with performance on certain types of problems. Finally, based on analysis of the prevalence of various answer choices, common misconceptions are identified.

INTRODUCTION

Assessment is increasingly recognized as crucial to improving learning [1,2]. Since conceptual understanding is strongly tied to the ability to transfer newly acquired knowledge [3] – a major goal of many engineering science courses – assessment ought to include conceptual knowledge. One approach, pioneered with the Force Concept Inventory [4], has been to assess conceptual understanding through multiple choice tests in which questions and the various answer choices are carefully constructed to measure knowledge of important concepts and identify common misconceptions. Concept inventories have recently been under development for a number of engineering subjects [5].

One such inventory, the Statics Concept Inventory (henceforth referred to here as “the inventory”), has been developed to assess conceptual understanding for the engineering course Statics [6]. Underlying the test was an extensive observational analysis of the products of novice problem solving and a formulation of those observations into a conceptual framework [7]. The inventory comprises 27 questions which focus on 8 concepts in Statics.

Like most concept inventories, this test is intended to be administered towards the beginning and towards the end of a Statics course, offering a summative assessment of conceptual gains. However, there is also the potential for using the inventory as a means of formative assessment, particularly since the analysis of the test results provides scores on distinct concepts. Consider, for example, the inventory to be administered a few weeks prior to the end of Statics,

or at the beginning of a follow-on course (e.g., Dynamics or Mechanics of Materials). Then, the performance of individual students on different concepts could be fed back in real time to enable additional instruction that is tailored to the specific needs of individual students.

The current paper presents results for the 2004-2005 version of the inventory, which has been administered at a number of institutions. In addition, through detailed comparisons of inventory results with performances on course examinations, evidence is offered for the potential value of such concept-specific feedback.

RESULTS FOR 2004-2005 ADMINISTRATIONS

This paper reports on data from administrations of the inventory during the 2004-2005 academic year at US universities. In all cases, the test was administered near or after the completion of Statics (post-test) to a total of 1331 students in 10 different classes (Table 1). (The test was administered at two additional schools; one lost the data and a second mistakenly used the 2003-2004 test.) As reported below, the inventory was also administered as a pre-test (prior to Statics) in six of these classes.

In some cases, the test was administered in class with pencil and paper; in other cases, students took the test electronically. (In one class, students entered answers into a spreadsheet which was emailed in.) Students were aware that their score on the inventory would not affect their class grade, although at some institutions points were given for merely completing the test. Time limits of 50 to 60 minutes were imposed for tests taken in class; no limit is imposed

when the test is completed electronically. While students were told to complete the test on their own, there was no monitoring of students who took the test electronically. As will be seen from the data below, no noticeable differences in performance were found between different methods of test administration.

Table 1. Classes taking inventory (Test method: Paper and Pencil: 1, 4, 5, 7, 10; Electronic: 2, 3, 6, 8, 9.)

Class	N	Course	University
1	97	Statics	A
2	38	Statics	B
3	35	Statics	B
4	42	Statics	C
5	419	Statics	D
6	262	Mech. Of Mat.	D
7	158	Statics	E
8	75	Mech. Of Mat.	E
9	42	Mech. Of Mat.	F
10	163	Statics	G

Classes 2 and 3 were taught by the same instructor in two successive semesters of Statics. Classes 5 and 6 correspond to Statics and Mechanics of Materials, respectively, taught during successive semesters at the same university; the Mechanics of Materials course consisted mostly of students from the prior Statics course. Classes 7 and 8 likewise corresponded to Statics and Mechanics of Materials courses. With the exception of Class 1, which is a private research university (CMU), the remaining classes were at public universities, all with graduate and undergraduate programs with various rankings.

Means and standard deviations for the post-test scores, and the mean for the pre-test scores (where administered), are shown in Table 2 for each class. In the transition from a paper-based to web-based administration, personal data was not collected for many students. Therefore, the authors cannot report reliably on any differences of performance between genders and ethnic groups.

With the exception of Class 1, all schools had highly comparable pre-test scores. Students in class 4 were instructed to leave blank any questions which they did not know how to answer. Considering the fraction of questions which were left blank, the pre-test scores in class 4 would likely have been comparable to other schools. It can be seen that inventory scores were generally similar for classes taking the paper and pencil and electronic formats of the inventory.

Table 2. Means and standard deviations for Post-test, means for pre-test, and normalized gains.

Class	Post-Test	S.D.	Pre-Test	Norm.Gain
1	20.51	4.24	10.46	0.61
2	13.87	4.66	7.04	0.34
3	14.17	4.12	7.13	0.35
4	11.62	4.14	4.82	0.31
5	12.14	4.59	7.35	0.24
6	12.08	4.92	*	*
7	13.64	5.14	7.41	0.32
8	14.37	5.09	*	*
9	12.24	4.10	*	*
10	13.04	4.96	*	*

The variation in the post-test scores for different classes is more significant. Normalized gain was used by Hake to compare pre-post differences in performances on the Force Concept Inventory among classes with very different starting points [8]. Normalized gain is defined as the improvement divided by the maximum possible improvement. Clearly, the use of normalized gain is less important here, where classes have similar pre-test scores.

Class 1 had a 3-week introduction to Statics in the prior year as part of a freshman Fundamentals of Mechanical Engineering course. The relatively high post-test scores in Class 1 should also be viewed in light of the fact that the instructor is also the main developer of the inventory (first author). As was the case for all other classes, students in Class 1 did not see the inventory questions during the course of the semester. Nevertheless, it is likely that instruction in this class was influenced by the developer/instructor's views of Statics. Still, the higher pre-test scores for this class may indicate a greater readiness for the subject which allowed for greater gains.

A more extensive discussion of the evidence supporting the reliability and validity of the 2003-2004 Statics Concept Inventory scores has been presented [5]. For the current version of the inventory, Cronbach's coefficient alpha, a measure of the internal consistency reliability, was found to be 0.82. This indicates that the inventory items consistently measure the same underlying content.

One measure of the quality of an individual test question is its ability to discriminate between students. A question is more likely to be deemed poor, if students who answered it correctly performed worse generally on other questions as compared with students who answered it incorrectly. To determine the extent to which individual questions discriminate

between students, the sample was divided into students who scored in the upper 27% overall and those who scored in the lower 27% overall. The fraction of students in these two groups that answers a given question correctly was then computed. The Discrimination Index for that question is defined as the difference between these fractions. The numbers of questions with Discrimination Index in various ranges are shown in Fig. 1. Questions with Discrimination Index above 0.4 are considered very good discriminators, between 0.3 and 0.4 to be good discriminators, and below 0.3 to be marginal or poor discriminators [9]. Most items on this test discriminate quite well.

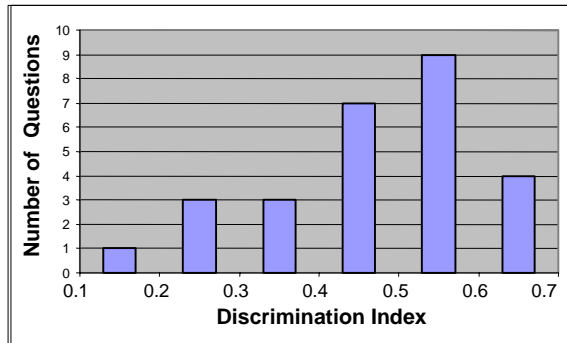


Fig. 1. Numbers of questions having Discrimination Index in various ranges.

RELATION TO CLASS EXAMINATIONS

Ultimately, instructors are concerned that students perform well in Statics courses, and also in subsequent courses. Thus, the inventory will be a more valuable tool, if it is indicative of performance on course examinations. Inventory scores of students were correlated with their examination scores (when they were available) using the Pearson Correlation coefficient r . For four Statics courses, the comparison was based on the final examination (Table 3); in Class 1, there was no final, so the correlation was based on the average of all 4 exams spread throughout the semester. For the two Mechanics of Materials courses, the inventory was compared with each of three course examinations (Table 4).

Table 3. Correlations between inventory and final exams in Statics classes.

Class	1	2	3	7	10
r	0.62	0.59	0.24	0.48	0.41

Table 4. Correlations between inventory and 1st, 2nd and final exams in Mechanics of Materials classes.

Class	8			9		
r	0.65	0.57	0.39	0.46	0.19	0.30

Given that the inventory asks questions on highly isolated and idealized aspects of statics, we argue that these correlations are quite meaningful. As a point of comparison, consider the correlations between exams *within each class* (Table 5). For example, in Class 1, the correlations 0.65, 0.66, 0.63, 0.59, 0.63, and 0.71 are between exams 1 and 2, 1 and 3, 1 and 4, 2 and 3, 2 and 4, and 3 and 4, respectively. With four exams, there are six pair-wise correlations; with three exams, there are three pair-wise correlations. By comparing the correlations provided in Tables 5 and 6, it can be seen that the correlations between the inventory and class exams are generally of the same order as correlations between class exams within the classes

Table 5. Correlations between course examinations, which are found to be comparable to correlations between inventory and examinations.

Class	Correlations between course examinations
1	0.65, 0.66, 0.63, 0.59, 0.63, 0.71
2	0.57, 0.34, 0.42, 0.55, 0.71, 0.73
3	0.42, 0.33, 0.66, 0.13, 0.25, 0.47
7	0.32, 0.44, 0.49, 0.48, 0.48, 0.59
8	0.60, 0.60, 0.70
9	0.54, 0.69, 0.64

CONCEPT-SPECIFIC SUB-SCORES

Feedback on students' performances that is more specific will be of the greater value to instructors than merely overall scores. The questions in the inventory were each devised to correspond to one of several concepts in Statics, as follows:

- Separating bodies (FBD): Identifying forces acting on a subset of a system of bodies
- Equilibrium (Equil.): consideration of both force and moment balance in equilibrium
- Friction: trade-off between implications of equilibrium and Coulomb's Law
- Static Equivalence (St.Eq.): Static equivalence between forces, couples and combinations
- Roller: Direction of force between a roller and the rolled surface
- Negligible Friction (Neg.Fr.): Direction of force between frictionless bodies in point contact
- Slot: Direction of force between a pin and the slot of a member
- Representation (Repres.): Representing unknown loads at connections

Hence, one should be able to infer from scores information regarding specific concepts and provide it to instructors. To investigate the feasibility and

usefulness of such inferences, we have undertaken: (1) to determine if independent sub-scores can be extracted consistently from the inventory; and (2) if concept-specific sub-scores are correlated with concept-specific performance in class examinations.

The psychometric analysis regarding sub-scores has included extensive factor analysis (both exploratory and confirmatory). Details of this analysis will be presented elsewhere. In short, exploratory factor analysis provided evidence that sub-scales arise from the data, and confirmatory factor analysis showed that these sub-scales are associated with the distinct concepts on which the test questions are based. In the following section we offer comparisons between performance on course examinations and specific concept sub-scores from the inventory.

EXAM ERRORS AND CONCEPT SUB-SCORES

Given that statistical analysis legitimized the extraction of concept specific sub-scores from the data, how are sub-scores related to performance in course examinations? Two examination problems from Class 1, depicted in Fig. 2 and 3, were used to study these relations in detail. From the problem in Fig. 2, we assessed the ability of students to recognize the directions of force exerted by a roller on a surface and by a pin on a slot. From the problem in Fig. 3, we assessed the ability of students to allow the friction force to be less than μN , if equilibrium is satisfied and no slippage occurs.

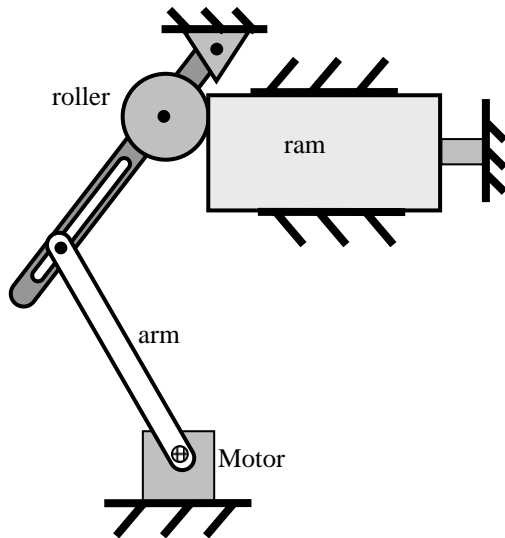


Fig. 2. Statics examination problem (Class 1) with multiple connected bodies (dimensions removed). The motor exerts a given torque on the arm, which leads the ram to crush the object with a force to be determined. Problem analyzed for errors in forces at

roller and slot (usual assumptions regarding frictionless pins apply).

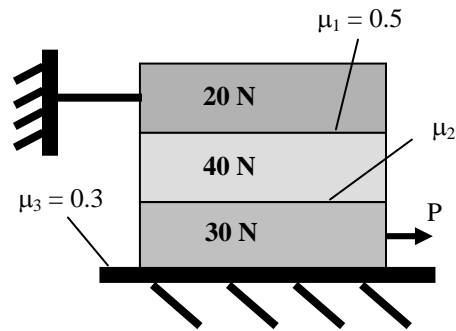


Fig. 3. Statics examination problem (Class 1) with friction. Levels of force P to cause sliding of the blocks to be determined for different levels of friction coefficient μ_2 . Problem analyzed for error of automatically associating the friction force with μN .

Students in Class 1 were divided into two groups: those who erred and did not err in recognizing the direction of the force of the roller in the exam problem of Fig. 2. Students in Class 1 were likewise split into those who did and did not err in the force of the pin on the slot (Fig. 2) and in the friction force (Fig. 3). For all three splittings of the class, students who erred had on average lower overall inventory scores and lower sub-scores on all concepts. However, for some concepts, the differences between the sub-scores of those who erred and did not err were not statistically significant.

On the other hand, one expects the sub-score Roller, for example, to be significantly different for the two groups who erred and did not err on the roller force in the examination problem. For each of the other exam errors, there is one concept sub-score (Slot and Friction) which also ought to discriminate most strongly between those who erred and those who did not. Based on their content, one expects the remaining concepts to be less directly related to the three exam errors.

In Table 6, we display the mean inventory sub-scores on these three concepts for the groups of students that did and did not make these three exam errors. In addition, a t-test was performed, and the p-value for the test that the two groups have the same mean concept sub-score is shown. The smallest p-values are for concept sub-scores that correspond to the concept at issue in the exam problem. This indicates that groups that erred and did not err in the use of a concept in a class exam differed significantly on those inventory questions specifically addressing that concept. In fact, with one exception (Roller sub-

score on the slot error), other differences between the groups were not statistically significant ($p > 0.05$).

Table 6. Comparison of concept sub-scores for students who made and did not make specific errors in course examinations. Most significant differences between groups were for concept most directly related to error.

		Roller	Slot	Friction
Roller: 47/97	Error	0.645	0.872	0.652
	No Error	0.867	0.927	0.760
Erred	p	0.002	0.199	0.127
Slot: 33/97	Error	0.646	0.778	0.667
	No Error	0.818	0.964	0.729
Erred	p	0.023	0.001	0.42
Friction: 53/97	Error	0.723	0.868	0.610
	No Error	0.803	0.939	0.826
Erred	p	0.265	0.075	0.001

The above exam-inventory comparison was based on results from Class 1 (students taught by the primary inventory developer). Given the correlations between class examinations and overall inventory scores shown above for other universities, it is likely that similar comparisons could be made for exam problems at other schools. One such analysis was conducted on a problem on the final Exam for Class 2 (Fig. 4).

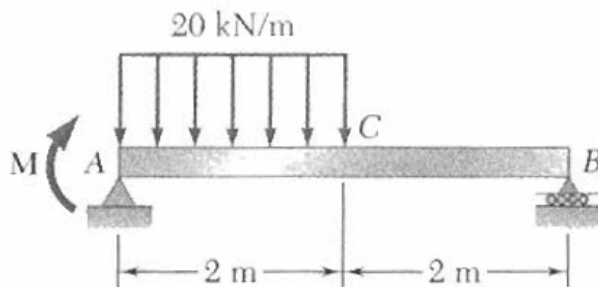


Fig. 4. Statics examination problem from Class 2 compared to inventory results. Students asked to draw the shear force and bending moment diagrams.

Grading in this problem was done by the instructor of Class 2 prior to, and independent of, comparisons with the inventory. Scores for this problem were based primarily on a student's ability to draw free body diagrams and write down equations of equilibrium, all of which lead to shear force and bending moment diagrams. Students could earn a maximum of 6 points, and one point was taken off for very minor errors. Thus, we compared inventory scores of two groups of students: those who earned 5 or 6 and those who earned less than 5, indicating more substantial errors in free body diagrams or equilibrium conditions (Table 7). Sub-scores on all

concepts are shown, along with (i) the difference in the means divided by the standard deviation of the sub-scores for the full group, and (ii) the p -value from a t -test comparing the means of the two groups.

Table 7. Inventory performances of students in Class 2 with at most minor errors (score 5 or 6, $N = 19$) on problem in Fig. 4, compared to students with more significant errors (score 4 or less, $N = 19$).

	Score 5-6	Score 0-4	Diff/SD	p
FBD	0.80	0.55	0.80	0.013
Equil.	0.58	0.34	0.90	0.004
Friction	0.33	0.18	0.56	0.087
St. Eq.	0.33	0.12	0.78	0.015
Roller	0.82	0.61	0.65	0.045
Neg. Fr.	0.46	0.32	0.53	0.102
Slot	0.77	0.67	0.36	0.278
Repres.	0.67	0.49	0.52	0.109
Total	16.47	11.26	1.12	<0.0005

One should note the concepts that most discriminate between students who did well (5-6) on this exam problem and those who did not (0-4). Sub-scores on concepts that are irrelevant to this problem, namely Friction, Neglect of Friction, Slot, and Representing Forces at Connections, display non-significant differences. (The concept Roller is also irrelevant to this problem; with $p = 0.045$, its difference is marginally significant.) On the other hand, the most relevant concepts, Free Body Diagrams, Equilibrium, and Static Equivalence, display the most significant differences. The relevance of these concepts to this problem is noteworthy. Certainly, the inventory was devised with a focus on the concepts necessary to solve problems of multiple inter-connected bodies [6], such as that of Fig. 2. It is clear, however, that the inventory is also quite relevant to problems that instructors would view as very distinct, namely, the drawing of shear force and bending moment diagrams, as demanded by the problem in Fig. 4.

COMMON MISCONCEPTIONS IN STATICS

One important use for the inventory is to assess the prevalence of various types of errors. These can point to widely held misconceptions, which ought to be addressed in instruction. Here we point out errors which occurred most frequently.

An example of a question in the group "FBD" is shown in Fig. 5. The student is asked to choose the correct free body diagram for the collection of blocks 1, 3 and 6, and the cords connecting them. The most commonly chosen wrong answer was the distractor

with an “internal force”, that is, a force acting between two of the bodies in the diagram (T_C in Fig. 6). This incorrect answer was chosen by 15% of students, while the correct answer was chosen by 66% of students. (In other questions, the internal force distractor was chosen by 26% of students).

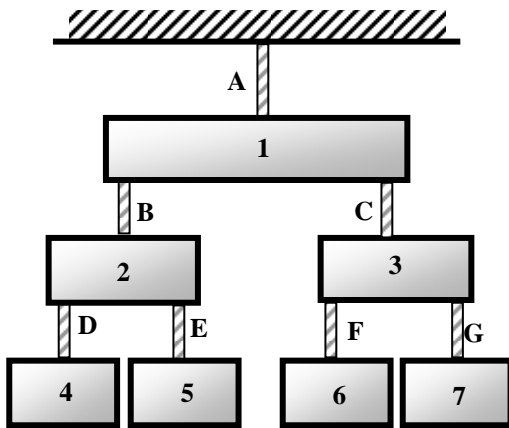


Fig. 5. Question requesting student to draw the free body diagram of a subset of a system.

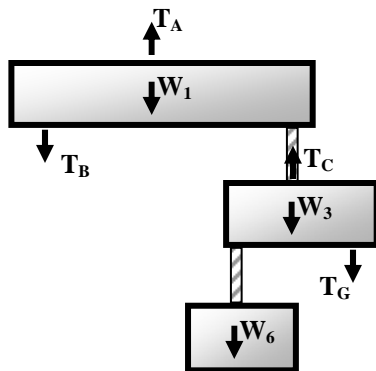


Fig. 6. Commonly selected wrong answer to question in Fig. 5, featuring an internal force.

An example of a question in the group “Static Equivalence” is shown in Fig. 7. Students are given one loading and asked to find a second loading that is equivalent. The most commonly chosen wrong answer was the distractor (Fig. 8) in which the couple is taken to be equivalent to a force that apparently produces the same moment. This incorrect answer was chosen by 26% of students, while the correct answer was chosen by 32% of students. For one question in the group “Static Equivalence”, distractors of this type were chosen more frequently than the correct answer.)

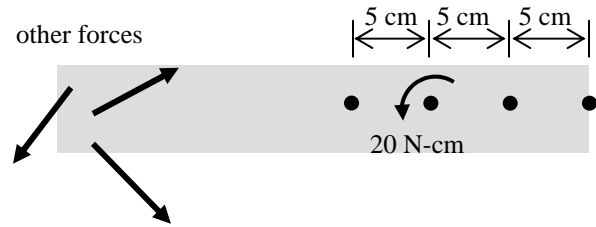


Fig. 7. Question requesting student to choose loading that is statically equivalent (to the 20 N-m couple).



Fig. 8. Commonly chosen answer to question in Fig. 7, (force taken to be statically equivalent to couple).

An example of a question in the group “Roller” is shown in Fig. 9. Students are shown a system and asked for the direction of the force of the roller on the body it rolls on (the usual assumptions of a frictionless pin and so forth are given). The most commonly chosen wrong answer was the distractor in which the force is acting parallel to the arm to which the roller is pinned, rather than perpendicularly to the rolled surface. This incorrect answer was chosen by 20% of students, while the correct answer was chosen by 58% of students. In general, there is a strong tendency to assume that forces always act parallel to elongated members (as they do in a two-force member).

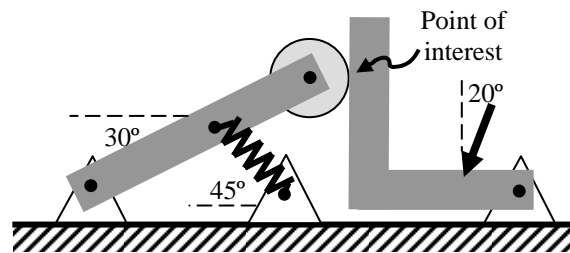


Fig. 9. Question requesting student to choose direction of force on surface contacted by roller.

An example of a question in the group “Friction” is shown in Fig. 10. In this question, students are asked for the upward force exerted by the left block on a center block. The choices are all numbers. The distractor of 8 N was chosen by 39% of students, while the correct answer of 3 N was chosen by only 32% of students. Students are powerfully drawn to assume that the friction force equals the coefficient of friction times the normal force, rather than choosing a lesser force that maintains equilibrium.

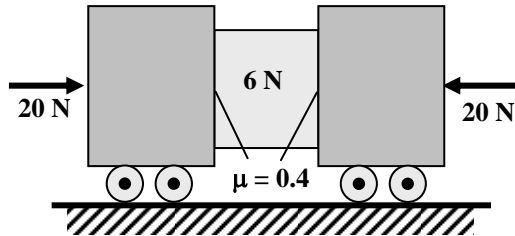


Fig. 10. Question requesting student to choose magnitude of force between bodies in frictional contact.

An example of question in the group “Equilibrium” is shown in Fig. 11. Students are shown the bar (a) with the force applied and asked for the loading exerted by a hand that grips the right end and maintains the bar in equilibrium. The most commonly chosen wrong answer is the distractor in which a force and a moment try to balance each other. The incorrect answer, shown as b in Fig. 11, was chosen by 25% of students, while the correct answer was chosen by 58% of students. This parallels the error in static equivalency.

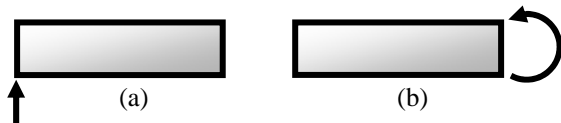


Fig. 11. (a) Question in which student is requested to find load exerted by hand at right which balances bar. (b) Most commonly chosen wrong answer.

A second example of a question in the group “Equilibrium” is shown in Fig. 12. The student is asked whether one or both of the loadings could be in equilibrium (all forces and couple have positive magnitudes and act in the direction shown, but the magnitudes are adjustable). Very few students (11%) recognized that neither loading could produce equilibrium, (I) because forces cannot balance and (II) because moments cannot balance. In particular, 70% of students accepted that (II) could be in equilibrium, while 53% of students accepted that (I) could be in equilibrium.

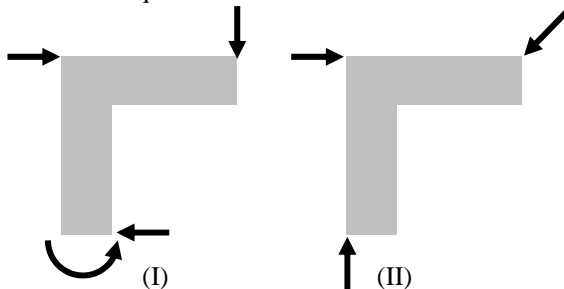


Fig. 12. Question in which student is asked if loadings (I) and/or (II) could be in equilibrium.

Students exhibit a wide range of misconceptions, as evidenced by the choice of wrong answers; however, certain errors stand out. Many students have trouble rejecting internal forces, perhaps revealing a lack of clarity regarding what body exerts each force or which bodies’ forces ought legitimately to be included in a free body diagram. Students often fail to grasp the necessity of independently satisfying force and moment summation and/or that a couple carries no net force. Forces at various connections sometimes have their directions set by the nature of the connection, whereas students are distracted from this by the shape of the body or by other applied forces. Finally, the quantity μN is only the limit on the friction force; students are too often convinced that μN must be the actual level of the friction force, even when there is no slip and a lesser force is sufficient for equilibrium.

SUMMARY AND CONCLUSIONS

The Statics Concept Inventory is a multiple choice test that assesses conceptual knowledge of students in Statics. The test consists of 27 questions that capture 8 distinct concepts and include distractors (wrong answers) that have been constructed based on observations of student work. This paper is based on results from the administration of this test to 1331 engineering students in 10 classes from 7 US universities during the 2004-2005 academic year. Previous observations as to the psychometric soundness of the test, including reliability and quality of items in terms of discrimination indices have been reconfirmed. The bulk of the paper addresses comparisons between performance on the inventory and performance on class examinations. It has been found that there are significant correlations between overall inventory scores and examination scores, correlations that are on the order of correlations between class examinations. In addition, we have shown that inventory sub-scores on specific concepts can offer insight into the propensity to make the analogous conceptual errors in examinations. Finally, by investigating the distribution of wrong answers, we have identified common misconceptions, information that may be valuable for instruction. The inventory is available for all instructors, who can learn more by visiting: www.engineering-education.com/CATS.

ACKNOWLEDGEMENTS

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

BIBLIOGRAPHIC INFORMATION

1. P. Black and D. William, Assessment and Classroom Learning, *Assessment in Education*, Vol. 5(1), pp. 7-73, (1998).
2. National Research Council, *Knowing What Students Know: the Science and Design of Educational Assessment*, J.W. Pellegrino, N. Chudowsky, and R. Glaser (Eds.), Washington, D.C., National Academy Press, (2001).
3. National Research Council, *How people learn: Brain, mind, experience and school*, Committee on Developments in the Science of Learning, Bransford, J.D., Brown, A.L., Cocking, R.R. (Eds.), Washington, D.C., National Academy Press, (1999).
4. D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory", *The Physics Teacher*, Vol. 30, p. 141-158, (1992).
5. D. Evans, et. al., Panel Discussion: Progress on Concept Inventory Assessment Tools. 33rd ASEE/IEEE Frontiers in Education Conference, Boulder, Co., November 5-8, (2003).
6. P.S. Steif and J.A. Dantzer, "A Statics Concept Inventory: Development and Psychometric Analysis" (to be published in *Journal of Engineering Education*).
7. P.S. Steif, An Articulation of the Concepts and Skills which Underlie Engineering Statics, 34th ASEE/IEEE Frontiers in Education Conference, Savannah, GA, October 20-23, (2004).
8. R. Hake, Interactive Engagement Versus Traditional Methods: a Six-Thousand Student Survey of Mechanics Test Data for Introductory Physics Courses, *American Journal of Physics*, 66, 1, pp. 64-74, (1998).
9. L. Crocker and J. Algina, *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston, Inc., Philadelphia, (1986).