2006-204: FEEDING BACK RESULTS FROM A STATICS CONCEPT INVENTORY TO IMPROVE INSTRUCTION

Paul Steif, Carnegie Mellon University

Professor, Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pa Degrees: Sc. B. 1979, Brown University; M.S. 1980, Ph.D. 1982, Harvard University. Research area: solid mechanics and engineering education.

Mary Hansen, Robert Morris University

Mary A. Hansen Assistant Professor, School of Education and Social Sciences, Robert Morris University, Moon Township, PA Degrees: B.S. 1994, California University of Pennsylvania; M.A., 1996, M.S. 1999, Ph.D., 2005, University of Pittsburgh. Research Area: Educational Assessment

Feeding Back Results from a Statics Concept Inventory to Improve Instruction

Introduction

Effective assessment is known to be critical to improving learning outcomes[1,2]. For many engineering subjects, one hopes students will learn to *transfer* their newly gained knowledge to new situations, which then requires a deep understanding of the material[3]. This has been taken to mean conceptual understanding. One approach to assessing conceptual understanding, with its origins in the science education community, is the Force Concept Inventory[4]. The approach of concept inventories has been extended by the engineering education community to a variety of engineering subjects[5]. If such concept inventories can provide formative assessment - feedback to improve the learning of students who take them - then their usefulness will be further enhanced.

In the present paper we consider the use of the Statics Concept Inventory (SCI) as a basis for formative assessment. This test has been reported on previously[6,7]. We must demonstrate that the results of this inventory are meaningful and worthwhile to feed back, and that we can devise effective formats for presenting these results. Accordingly, we report briefly on indicators of the quality of the test results. Next, we show how results can be presented to help instructors compare their students' performance with those of other institutions and to help students compare their performance with those of peers in their class. Finally, we present survey data from students who attended a review session addressing SCI questions prior to the class final exam.

Background on Statics Concept Inventory and Current Status

The SCI consists of 27 multiple choice questions; each question addresses a single concept, involves negligible calculation, and features wrong answers that capture typical conceptual errors made by students. The conceptual framework and typical errors were based on field studies of students' work as described elsewhere[8]. Psychometric analyses of the test results as a whole and its individual items have been conducted. Such analyses, which draw on statistical methods, are aimed at judging whether the test yields measurements of knowledge or skills that are reliable and valid. Based on such psychometric analyses, the test has steadily improved each year. The numbers of students taking the test has also increased: 245 in 2003-2004, 1330 in 2004-2005, and 1255 as of the first half of 2005-2006 (with 16 classes participating). Only very minor changes to the current version are anticipated for the future.

A critical feature of the test is that questions are grouped according to concept. In the 2005-2006 version of the test, there are 9 concepts, with 3 questions per concept. The concepts are given in Table 1.

Table 1. Concepts tested in 2005-2006 version of Statics Concept Inventory.

Concept	Description
A	Drawing forces on separated bodies
B	Newton's 3rd Law
C	Static equivalence
D	Roller joint
E	Pin-in-slot joint
F	Loads at surfaces with negligible friction
G	Representing loads at connections
H	Limits on friction force
I	Equilibrium

Grouping questions according to concept is critical because students and instructors cannot take specific actions based on the total score. However, if performance in each of a set of concepts is fed back, then students have more guidance for their subsequent study efforts.

Observations of Typical Errors

We summarize here a recently reported evaluation of the most widely held misconceptions, as judged by the relative prevalence of wrong answer choices in the inventory. When confronted with drawing a free body diagram of a collection of bodies, many students include internal forces; i.e., forces between bodies included in the diagram. This reveals perhaps a lack of clarity regarding what body exerts each force or which bodies' forces ought legitimately to be included in a free body diagram. As is evident from questions on Static equivalence, students often take a couple to be equivalent to a force which produces the same moment about some point. The concepts of force, moment and couple do not seem to be adequately distinguished. In equilibrium, students often fail to impose both equilibrium of forces and of moments. Forces at some connections, e.g, rollers, pins in slots, have their directions set by the nature of the connection. Students allow themselves to be misled by the presence of applied loads, or by the shape of the body, into thinking the force direction is other than the connection would allow. (Superficial application of equilibrium considerations on incompletely defined bodies may convince students that the force direction is influenced by the applied load. 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.

Psychometric Indicators of Overall Test Results

Unless otherwise indicated, the measures presented here pertain to data obtained in the fall 2005 semester based on 1164 students in 16 classes (at 14 different institutions) who took the web version of the test. Data from students who spent less than 10 minutes on the test were excluded from the original sample size of 1255 (scores of students taking less than 5 minutes were nearly equivalent to guessing and those who spent between 5 and 10 minutes were only marginally better).

Cronbach's alpha measures the internal consistency or correlation between questions, also referred to as "items". In simple terms, students who do well on one question should be more likely to do well on others. Cronbach's α is defined in terms of r_m, the mean of the inter-item correlations, and N the number of items according to:

$$\alpha = \frac{N r_m}{1 + (N-1) r_m} \qquad (1)$$

If all questions capture exactly the same overall knowledge construct, say "Statics Concept Knowledge", there would be a perfect correlation between all the items $r_m = 1$; then α would be 1. For the test as a whole $\alpha = 0.834$. Values in excess of 0.8 are considered as indicating good internal consistency.

The discrimination index is one measure of the quality of an individual item or question. Item discrimination refers to the ability of an item to distinguish between examinees of higher ability (those who do well on the test) and lower ability (those who do poorly on the test). We compare two groups: students whose overall scores were in the top 27% and those in the bottom 27%. The difference in the fractions of students in these two groups who answer an individual item correctly is the discrimination index. One desires the discrimination index to be high: students who have better overall knowledge should be more likely to answer any given item correctly. Discrimination indices below 0.2 are considered to be poor, in excess of 0.3 to be good, and above 0.4 to be very good. In Figure 1, we show histograms of the numbers of questions that have different levels of discrimination index; the left histogram is for the 2004-2005 test, and the right is for the 2005-2006 test. While the indices are quite high for both years, there is a noticeable improvement from the past year to the current year. There is now only a single item with a discrimination value below .3 (the item answered correctly by very few examinees) and only three additional items with discrimination indices below .4. Thus, most of the items on the SCI individually appear to distinguish well between examinees of different knowledge levels.

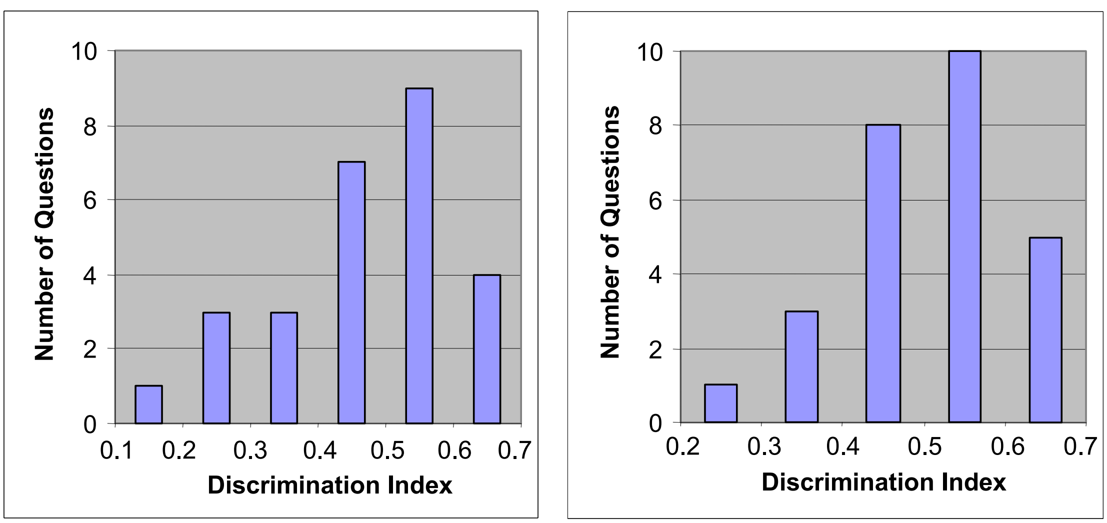

Figure 1. Histograms of discrimination indices for 2004-2005 and 2005-2006 versions of the Statics Concept Inventory.

Psychometric Indicators of Meaningfulness of Concept-based Sub-Scores

Since our goal is to provide actionable feedback to students and instructors, we believe that reporting only on the total score is insufficient, while reporting scores on individual items may be too fine-grained. Reporting performance on each of the 9 individual concepts – the mean over the three items comprising each concept – seems to be an appropriate level of detail for feedback. However, we first need to establish that such concept-level sub-scores are meaningful.

To be meaningful, the items that constitute a concept sub-score ought to be relatively well correlated with one another. That is, if a student answers one question correctly within the concept of, say, static equivalence, that student should be reasonably likely to have answered another question on static equivalence correctly. By contrast, whether the student answers a question on static equivalence correctly should have less bearing on whether a question on friction is answered correctly. To make this comparison, we calculated the full set of correlations between items, and then compared correlations between items testing the same concept with correlations between items testing different concepts. The appropriate correlation for dichotomous data (right or wrong, 1 or 0) is not the Pearson correlation, but the tetrachoric correlation[9].

In Table 2, we compare correlations within and across concepts. In all cases, the correlations between items within a concept are higher than those across concepts. In some instances, the correlations within a concept are significantly higher. Often, a relatively low mean correlation within a concept can result from one of the three questions not correlating with the other two questions; a single such outlier question is responsible for the low means in concept G and I. In general, the concept sub-scores appear to measure distinct types of knowledge, although clearly these types of knowledge are related, in that the correlations across concepts are not zero.

Table 2. Mean correlations between items within a concept and between items across different concepts.

	Overall	A	B	C	D	E	F	G	H	I
Within	0.500	0.672	0.416	0.466	0.663	0.672	0.406	0.258	0.613	0.333
Across	0.239	0.288	0.205	0.240	0.179	0.256	0.281	0.176	0.228	0.251

While not reported here, other studies of the SCI have found that its results correlate well with class exams at a number of institutions, and that performance on specific concept sub-scores correlate with the ability to use those specific concepts in exams[7]. The latter offers further support for the notion that concept sub-scores are meaningful.

Providing Feedback to Instructors and Students

If the scores on individual concepts are deemed meaningful, then providing them to instructors and students would seem to be valuable. Instructors might be interested in how their students performed, both relative to each other and relative to other institutions. Individual students might be particularly interested in how they faired relative to their classmates. Here we show several methods of display.

To assess the distribution in total scores, and comparisons with other institutions, we can provide instructors with boxplots such as shown in Figure 2. The box of a boxplot displays the 25th, 50th (median), and 75th percentile. The whiskers run to the 10th and 90th percentiles. Dots are outliers.

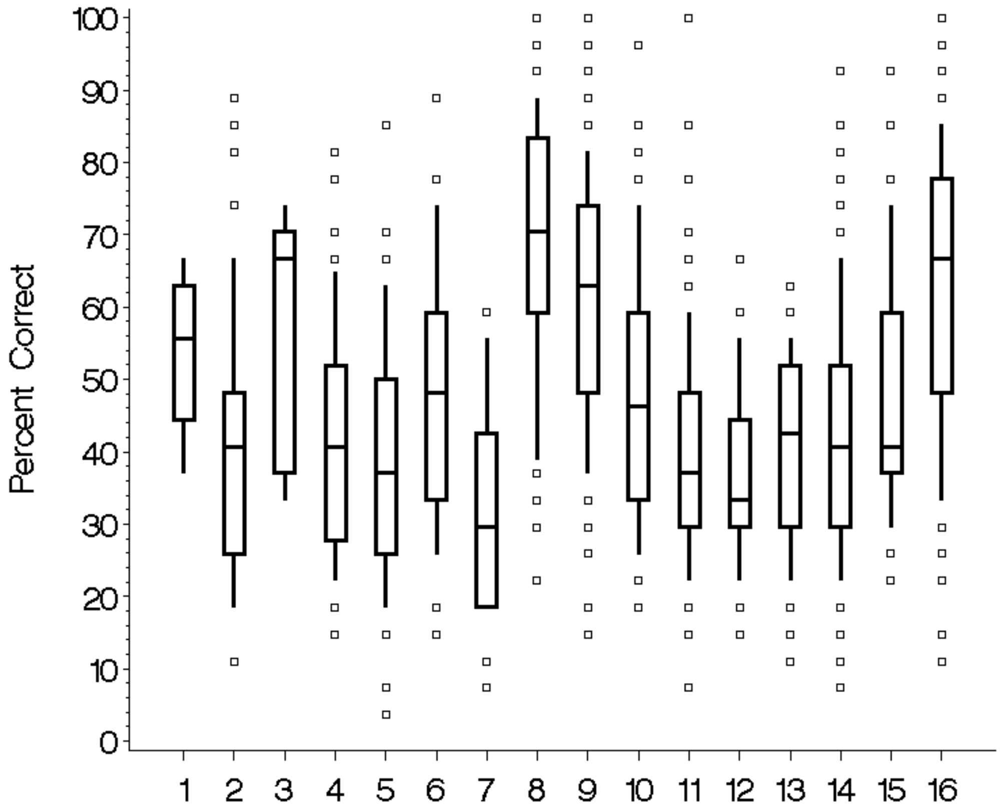

Figure 2. Boxplots for total SCI scores of 16 classes taking Statics Concept Inventory.

To allow concept-specific comparisons with other institutions, plots such as the one in Figure 3 are useful. The red dots denote the mean total and concept scores for Class #16 (red dots) compared to other classes (shown as black dots). Such results are displayed on our web site (http://engineering-education.com/CATS/results.htm) where viewers can examine similar plots for all 16 classes.

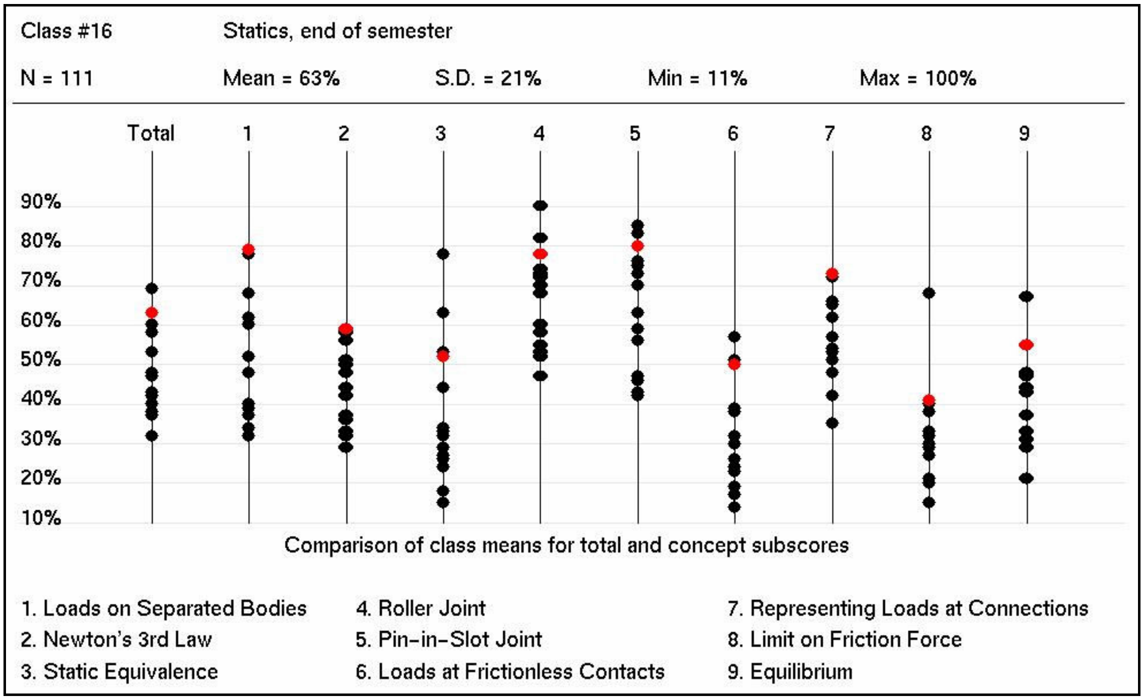

Figure 3. Comparisons of total and concept sub-sores between various classes taking Statics Concept Inventory

Finally, to allow students to judge their scores on various concepts relative to their peers, bar graphs such as that shown in Figure 4 can be prepared for each class. Students who take the test on the web are emailed back their scores automatically (individual items, concept scores and total). After the whole class has completed the test, such a graph can help students judge the concepts (and respective items) on which they should focus further attention.

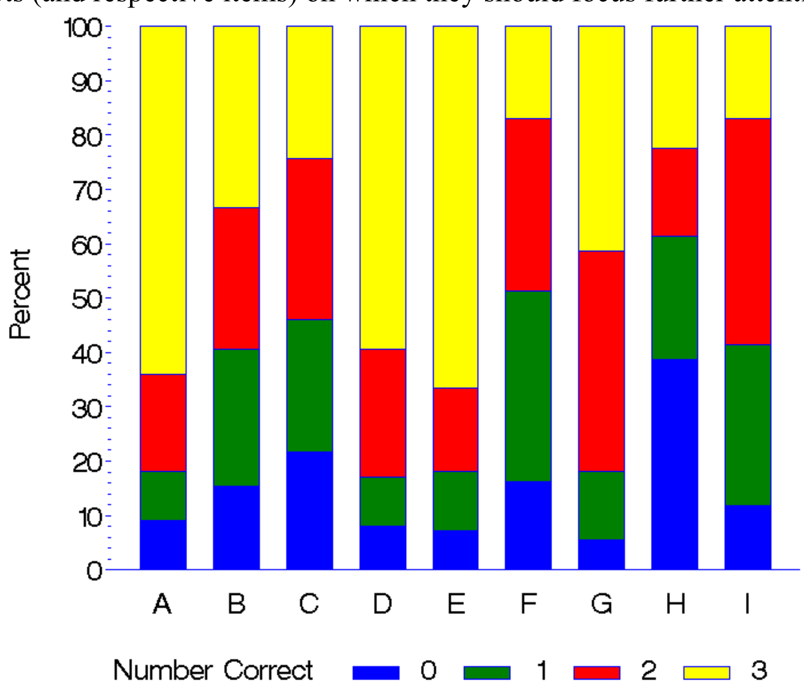

Figure 4. Percentage students answering questions correctly in concepts A to I. (Class 16).

Remedial Instruction to Address Conceptual Deficiencies

A review session for students was held by the instructor in Class 16 prior to the final exam. Students had already completed the SCI and received an email that listed their total score, score on each concept and whether each questions was wrong or right. The instructor had received results for the class as a whole listing performance on each concept and question. Based on that feedback, the instructor decided to focus discussion on the questions in the five concepts in which students had performed most poorly. Upon completion of the session, students were asked to complete a survey with the following questions:

1. Did this review session help you develop a better understanding of the concepts underlying the questions on this test?
2. Did you get more from this review session because you knew which questions and concepts you answered correctly and incorrectly on the test?
3. Do you expect that this review session will help you do better on the final exam?

A five-point Likert scale was used for responses, with the following labels: 1 (not at all), 3 (somewhat), and 5 (very much). Means and standard deviations for the 41 responses to the survey are shown in Table 3.

Table 3. Survey responses of students participating in review session in Class 16.

	Question 1	Question 2	Question 3
Mean	4.15	3.65	3.64
St. Dev.	0.70	1.03	0.90

Besides the generally favorable response evident in the data, student written comments were also nearly all positive. Recognizing its importance, two students pointed out that instructors should remind students to bring the emailed results of the SCI to the review session. The instructor also believed that the experience was worthwhile. It is critical to make the most of such opportunities for instruction tailored to observations.

Summary and Conclusions

This paper has reported on Statics Concept Inventory (SCI) results obtained during the Fall 2005 semester. The SCI measures a student's ability to use the key concepts of Statics individually and in isolation, with negligible mathematical analysis. This test includes 27 multiple choice questions, focusing on 9 concept areas, including: free body diagrams, relations between forces, couples and moments, the forces acting at connections and between bodies, and the conditions of equilibrium.

The SCI has been previously subject to extensive psychometric analyses, which have established high levels of reliability and validity. Here we give some indicators of the improvements in properties of the test over the past year. Further, we show that correlations between questions addressing the same concept are higher than correlations between questions across concepts; this points to the significance of sub-scores on the individual concept areas. It is argued that feedback to instructors and students that is more detailed than the total test score is called for, and that concept sub-scores may be feedback with appropriate granularity.

Graphical displays of the relevant results were then shown. Some displays provide insight to instructors as to how their students perform relative to students at other institutions. Other displays help students see how they perform relative to their classmates. Finally, we have reported on one class which held a review session to discuss questions from the SCI after students had been fed back results of their performance on the SCI. The response of students to this session was positive. Goals of our future efforts will be to refine the feedback we offer, to measure the effectiveness of such follow-on instruction, and to improve upon it.

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. Hestenes, D., Wells, M. and Swackhamer, G., "Force Concept Inventory", The Physics Teacher, Vol. 30, p. 141, 1992.
5. D. Evans, C. Midkiff, R. Miller, J. Morgan, S. Krause, J. Martin, B. Notaros, D. Rancor, and K. Wage, "Tools for Assessing Conceptual Understanding in the Engineering Sciences," Proceedings of the 2002 FIE Conference, Boston, MA.
6. Steif, P.S. and J.A. Dantzler, "A Statics Concept Inventory: Development and Psychometric Analysis", Journal of Engineering Education, J. Eng. Educ., Vol. 33, pp. 363-371, 2005.
7. Steif, P.S. and M.A. Hansen, "Comparisons Between Performances In A Statics Concept Inventory And Course Examinations", to be published in Int. J. Eng. Educ., 2006.
8. Steif, P.S., "An Articulation of the Concepts and Skills which Underlie Engineering Statics," 34rd ASEE/IEEE Frontiers in Education Conference, Savannah, GA, October 20-23, 2004.
9. Glass, G.V. & Hopkins, K.D. Statistical methods in education and psychology, Third Edition. Boston: Allyn and Bacon, Inc., 1996