2006-1262: SYSTEMATIC INTEGRATION OF CONCEPT INVENTORIES IN MECHANICAL ENGINEERING

Daria Kotys-Schwartz, University of Colorado-Boulder
DARIA KOTYS-SCHWARTZ is a doctoral candidate and instructor in the Department of Mechanical Engineering at the University of Colorado at Boulder. She earned B.S. and M.S. degrees in mechanical engineering at The Ohio State University. Her research interests include polymer processing, development of student assessment measures, gender disparity in engineering and innovative instructional methodology.

Derek Reamon, University of Colorado
DEREK REAMON is a senior instructor of mechanical engineering at the University of Colorado at Boulder, where he has taught Circuits and Electronics, Mechatronics, Component Design and the interdisciplinary First-Year Engineering Projects. He received his M.S. and Ph.D. in Mechanical Engineering from Stanford University. His foremost research interests include assessment of student learning, curriculum development and robotic controls.

Lawrence Carlson, University of Colorado-Boulder
LAWRENCE E. CARLSON is a founding co-director of the Integrated Teaching and Learning Laboratory and Program, as well as professor of mechanical engineering. He received his M.S. and D.Eng. degrees from the University of California at Berkeley. His primary educational passion is real-world design, and he spent his last sabbatical leave at IDEO in Palo Alto, CA, sharpening some rusty design tools.
Systematic Integration of Concept Inventories
In Mechanical Engineering

Abstract

Gauging student knowledge accurately is a complex task that has always challenged educators. The commonly employed metrics (homework, quizzes and exams) serve as indicators of student performance for instructors — yet are these instruments truly assessing student knowledge gains? Concept inventories have recently emerged as tools for assessing students’ understanding of the basic concepts upon which technical education is based. Initially developed to test learning of basic physics concepts, concept inventories have subsequently been developed and validated for a variety of engineering subjects. By undergoing a rigorous process of validation, engineering concept inventories can provide meaningful primary assessment throughout a curriculum. This, in turn, allows a methodical evaluation of the effectiveness of various teaching methods, enabling subsequent improvements in learning. However, concept inventories have not been applied in a systematic way to engineering curricula. This paper focuses on the systematic integration of eight previously developed and disparate concept inventories, utilized to assess the major portion of a mechanical engineering undergraduate curriculum. We present the preliminary results, and discuss the forthcoming efforts to develop and validate two additional concept inventories that will provide a full assessment package for the core mechanical engineering undergraduate curriculum.

Introduction

In engineering, faculty have been extensively trained in technical theory, but typically not in pedagogical skills or educational research endeavors that focus on student learning. Additionally, typical university reward structures do not encourage such a focus. As a result, many engineering faculty members resort to personal educational experiences as the primary resource for their teaching philosophy. In other words, they tend to teach the same way they were taught.

As engineering educators, we each have our own perception of the content knowledge that students gain in the classroom. However, it is often astonishing for an instructor to discover the difference between what students are actually learning and what an instructor perceives they are learning. It can be an extremely disheartening experience for an instructor to realize that at the end of term students still do not comprehend conceptual topics that were proficiently taught. It is essential for educators to acknowledge that students bring to the classroom experiences, attitudes and perceptions that can greatly influence classroom performance, and instructional modifications may be needed to address these student issues. The compelling questions remain: what misconceptions do students bring into the classroom, and what knowledge and understanding do they depart the course with?

Throughout the United States, engineering programs are faced with the daunting task of developing specific methods and assessment tools that meet the requirements of the current ABET Engineering Criteria. The intent of these recently established ABET criteria is that
engineering programs will implement assessment instruments that encourage continuous improvement. Consequently, engineering programs are increasingly recognizing the need to employ quantitative assessment measures to evaluate both the “hard” engineering skills and the “professional” skills. Currently, all engineering programs must implement an on-going evaluation system that demonstrates that their students have attained program-specified objectives and outcomes. However, there are limited reliable assessment instruments, based on rigorous research methods, available for implementation, even for the “hard” skills.

For instance, in our mechanical engineering (ME) department, many courses in the curriculum are prerequisites for advanced required courses, and students must earn a C- grade (or better) in order to take the subsequent course. Proficiency exams were developed in-house beginning in 1999 to test student learning in prerequisite courses. The goal was to not only provide valuable assessment information on the prerequisite course, but to also give the instructor of the next course a clear understanding of his/her students’ knowledge of prerequisite material — to allow better tailoring of course material, thus improving both courses. The exams were designed to be assigned as homework during the first week of the subsequent course; they were implemented and administered on-line using WebCT.

A number of difficulties have been encountered with the ME proficiency exams. First, the exams were typically written by individual faculty members in isolation, so the exams lack consistency of length, difficulty and concept/computational balance, making the results difficult to compare between courses. Second, not all faculty of the subsequent courses “bought in” to the administration of the exams. Every semester, proficiency exam results were sent to the specific faculty involved with the pre- and post requisite courses. However, the data was not incorporated into the overall curriculum assessment process in an organized way, although that was the original intent. Additionally, the pedagogical purpose of the exams was varied, and faculty disagreed over their utility. As a result, some faculty assigned no weight to the exams in their final grades, so students had no incentive to complete the exam, or faculty would assign a due date halfway through the semester, reducing the utility of the exam. Third, the position of the exams in the post requisite course was problematic. Students progressed through the curriculum at varying speeds, and not all of the students in the post requisite course had the same prerequisite experience, so their value as a timely assessment of a given course was diluted. Finally, the WebCT interface is not optimal for department assessment due to ongoing design and administrative issues. On-line administration also necessitates an open-book format, again impeding faculty buy-in.

Background

Force Concept Inventory (FCI)

During the mid-1980’s it was recognized by the physics community that students were leaving Newtonian physics courses with the ability to solve quantitative exercises, but often not the ability to answer a simple qualitative question regarding the underlying concept. This suggested that students were forming erroneous common-sense beliefs, or mental models, about the fundamental concepts underlying the mathematics. As a result, the Force Concept Inventory (FCI) was developed by Hestenes and Halloun. This multiple-choice exam was designed to assess student understanding of fundamental Newtonian physics, and identify student misconceptions. The results showed that some students who performed well on homework and exams were unable to transfer that knowledge to solve problems different than those shown in
class, yet founded upon the same topic area. The FCI empirically demonstrated that students lacked the conceptual understanding to correctly answer questions similar to those presented in class, but in a different form. Subsequent research has revealed that gains in student understanding of fundamental Newtonian physics could be achieved by employing alternative teaching techniques in the collegiate physics classroom. In a 6000-student FCI study, Hake demonstrated that students in a traditionally instructed course were not able to attain normalized FCI gains as high as students that participated in interactive engagement classes. Essentially, the students taught using a traditional lecture mode that achieved the highest normalized gains matched the performance of the lowest performing students that were taught using an interactive engagement classroom practice. During the last fifteen years, the FCI has become highly regarded in the physics education community and credited with stimulating instructional reform. The FCI is currently being utilized in high school and collegiate physics classrooms across the country.

At the University of Colorado, the very active Physics Education Research (PER) group has extensively used FCIs as a powerful quantitative assessment measure to perform rigorous educational research in the physics department. For the last three years, PER has regarded the FCI as an opportunity for faculty development — a tool for refocusing on fundamental concepts and an identification of what gains in conceptual knowledge were not being accomplished in the classroom. Some of the educational research being undertaken by the PER group includes assessment (conceptual, belief-oriented and epistemological), examination of successful educational reforms and replication studies, social and contextual foundations of student learning, and student problem-solving and technology in physics.

**Engineering Concept Inventories**

Inspired by the physics Force Concept Inventory, the Foundation Coalition — a partnership of six universities funded by the National Science Foundation — has collaborated to develop concept inventories (CIs) for specific engineering disciplines, including: Strength of Materials, Dynamics, Circuits, Fluid Mechanics, Heat Transfer, Thermodynamics and Materials. The primary objectives of the Foundation Coalition CIs are to assess student knowledge of fundamental concepts in engineering and to promote evaluation of teaching methodology. The inadequate supply of proven assessment measures has been cited by the Foundation Coalition members as “one of the hindrances to reform in science, technology, engineering and mathematics (STEM) education.” Engineering concept inventories have the potential to be powerful, reliable and valid assessment tools that drive instructional innovations in engineering education. The development of tested, reliable concept inventory exams, however, is a lengthy and time-intensive process.

Each subject developer has successfully created initial and revised versions of the concept inventories, and each exam is currently undergoing reliability and validity studies. A benefit of these invaluable assessment instruments is that they have the potential for continuous improvement and integrate suggestions from a community of colleagues in the development of new versions.
Some of the challenges experienced by Foundation Coalition concept inventory developers included the need for experienced faculty to determine fundamental concepts for each exam. There were also difficulties determining a balance in fundamental concept questions, and balancing the number of questions on prerequisite material versus questions being used to survey current concepts. The complexity associated with developing the exam, which has been experienced by every CI developer, is primarily due to the extensive testing required for adequate validation.\(^{17}\)

With the support of NSF, Paul Steif of Carnegie Mellon University has developed a Statics Concept Inventory which was created with the same vision and philosophy as the Foundation Coalition concept inventories, and encountered many of the same difficulties during development.\(^{18}\) Professor Steif has created initial and revised versions of the Statics Concept Inventory, and the exam is being subjected to on-going reliability and validity studies.

**Implementation of Concept Inventories**

From past experience with proficiency exams, we understood the difficulty with the administration of department-wide assessment programs which attempt to measure the effectiveness of required courses. On-going assessment will not be successful if faculty does not “buy-in” to the program, if students do not consider the exam a meaningful practice, if exam solutions are circulated among students, or if there is variability in the test administration. Success necessitates the full commitment of faculty, honest and consistent rewards for student participation, rigorous faculty training and strict exam administration to reduce variability.

Therefore, the first step for a successful assessment program is the commitment of the faculty. On May 4\(^{th}\), 2005 the current state of the ME proficiency exams, concept inventory background and proposed concept inventory implementation plan was presented to the mechanical engineering faculty. The concept inventory plan was unanimously and enthusiastically endorsed by the faculty present (80% of the department).

**Concept Inventory Acquisition and Adaptation**

Building upon the extensive engineering concept inventory research conducted by the NSF Foundation Coalition and Paul Steif of Carnegie Mellon, we mapped existing CIs to eight core ME required courses. The existing concept inventories that align with the core mechanical engineering curriculum are: Statics, Mechanics of Solids, Materials Science, Thermodynamics, Fluid Mechanics, Dynamics, Circuits and Heat Transfer. In the spring of 2005, each of the exams were requested from the author with the assurance that the exam integrity would be protected, and the results would be tracked and reported to assist with on-going reliability and validity studies.

The implementation of these proven evaluation methods began during the 2005-06 academic year, serving as a primary assessment tool for the mechanical engineering core curriculum. Before the fall 2005 semester commenced, ME faculty reviewed the CI exam specific to their core course, mapping the exam to the course learning objectives. Copies of each reviewed CI, exam instructions and bubble scoring sheets were made one week prior to the start of term. During this time pre-exam administration times were also scheduled by the ME department.
assessment coordinators, allowing for delivery of exam materials and clarification of administration procedures.

Administration Procedure

To reduce variability and maintain the reliability of the testing instrument, it is imperative that uniform procedures be used by faculty members administering the concept inventories. Detailed procedures for administrating the concept inventories and instructions for interpreting the results were documented and distributed to all mechanical engineering faculty members who teach a course in the core curriculum. The following policies have been incorporated into the concept inventory administration plan:

- Pre-exams are to be administered during first or second class meeting.
- Pre-exam reports are to be delivered to faculty members within the first two weeks of the semester.
- Post-exams are to be administered during the last class meeting (or during the last week).
- Post-exam reports are to be delivered to faculty members before the start of the subsequent term.
- Students may not use any electronic device, notes or text books during the concept inventory exam.
- All parts of the pre-post exams are numbered and should be collected to protect their security and integrity.
- Student participation (pre- and post-) should be encouraged by awarding a meaningful course grade for taking the exams. Suggested allocation is one homework grade for participation in both the pre- and post-exam.

In an effort to support the transition from proficiency exams to concept inventories, ME assessment coordinators provided assistance administering the pre-course concept inventory to ensure that unified and consistent procedures were taking place.

Developing Faculty Expertise

Utilizing concept inventories in a pre- and post-test manner can provide a wealth of informative feedback about students’ performance in a course (see Figure 1). From the pre-test, the instructor of Course I in a two-course sequence will learn what misconceptions students possess entering the course. This allows that instructor to tailor the material coverage accordingly, emphasizing areas of low understanding, and spending less time on the concepts students already have a good grasp of. The post-test results provide the Course I instructor with quantitative data on the student learning that has transpired during the course (or not!), thus assessing both the students and the course. The post-concept inventory results can also be used to inform the Course II instructor of incoming students’ knowledge of prerequisite material.

Taken together, the pre-post test sequence provides valuable information that can be used to improve the course. This aligns perfectly with ABET guidelines for a comprehensive plan for continual evaluation and improvement.
Initial Results

Students at other universities and colleges who have taken the Foundation Coalition CI exams generally experience normalized learning gains of 10-20% from pre- to post-exam. This aligns with the 10-20% increase typically found when administering the Physics Force Concept Inventory exam in a traditional lecture format course.

The results for the first semester CI exam implementation at our institution are presented in Table 1. Only 5 of the 8 core courses scheduled for concept inventory adaptation were offered during the fall of 2005. The Circuits course consisted of 77 students, with 70 students completing a pre- and post-concept inventory. In Dynamics, 108 of 152 students submitted pre- and post-CIs, while in Fluid Mechanics 122 out of 156 participants completed the two-part assessment. Within the Thermodynamics course a pre- and post-sample size of 112 was taken from a 130 total students. In Statics 124 students participated in the pre-concept inventory.

The strongest CI exam percentage gain we observed was in Thermodynamics, with students reporting a 22% concept inventory increase (42% change) across the 15-week course. The Circuits course saw scores increase 12% (30% change), with Fluid Mechanics and Dynamics courses experiencing gains of 15% and 6%, respectively (changes of 44% and 15%).

It should be noted that the Dynamics and Fluids Mechanics courses were taught using traditional lecture, and the Circuits and Thermodynamics courses integrated active learning techniques throughout the semester.

The Cronbach coefficient alpha has been used by Foundation Coalition members to evaluate the reliability of concept inventories. The coefficient alpha is a measure of the internal reliability of an assessment metric — or rather, the ability of the exam to evaluate the student consistently. The coefficient alpha can range from 0 to 1, with 0.7 or higher indicating the exam is reasonably reliable.
Looking at Table 1 it can be observed that Dynamics, Fluid Mechanics and Thermodynamic have alpha coefficients at 0.7 or higher. However, the Circuits exam has an alpha coefficient of 0.47. The authors believe that this is a result of modifying the Circuits CI exam before administration. The mechanical engineering Circuits Course Objectives incorporates elements from both circuit and electronics theory. As a result, questions were taken from both the Foundation Coalition Circuits Concept Inventory and the Electronics Concept Inventory. It is apparent that this action altered the validity of the exam, making the result for the course potentially erroneous.

The Statics post-CI was not administered due to scheduling conflicts and will be addressed in the Lessons Learned section. The Mechanics of Solids, Materials Science and Heat Transfer CIs will be implemented for the first time during the Spring 2006 semester due to the core curriculum schedule.

**Faculty Concept Inventory Results**

For the fall 2005 semester, the pre-CI results were reported to the faculty members during the second week of the semester. Utilizing Remark 5.5 software, test statistics, respondent statistics, grade frequency distribution, item/problem statistics and individual item/problem analysis were provided in a pre-course report. Looking at Figure 2, one can view a sample of the test statistics and item/problem analysis from a typical report provided to instructors. The test statistics provide a snapshot of the CI exam results — quantifying scores, standard deviation, variance and the Cronbach coefficient alpha. The item/problem analysis illustrates the student answer distribution for each exam problem in both table and chart form. This report section allows faculty to quickly identify concept areas where students possess misconceptions, and what specific misconceptions they are carrying with them to the course. An identical post-report was provided to faculty members during winter break, before the start of spring semester.

---

**Table 1: Fall 2005 ME Concept Inventory Results**

<table>
<thead>
<tr>
<th>Course</th>
<th>Pre-Course CI Exam Average</th>
<th>Post-Course CI Exam Average</th>
<th>Gain</th>
<th>Change</th>
<th>Alpha Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuits</td>
<td>40%</td>
<td>52%</td>
<td>+12%</td>
<td>+30%</td>
<td>0.47</td>
</tr>
<tr>
<td>Dynamics</td>
<td>39%</td>
<td>45%</td>
<td>+6%</td>
<td>+15%</td>
<td>0.76</td>
</tr>
<tr>
<td>Fluid Mechanics</td>
<td>34%</td>
<td>49%</td>
<td>+15%</td>
<td>+44%</td>
<td>0.70</td>
</tr>
<tr>
<td>Statics</td>
<td>20%</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Thermodynamics</td>
<td>53%</td>
<td>75%</td>
<td>+22%</td>
<td>+42%</td>
<td>0.73</td>
</tr>
</tbody>
</table>
It should be noted that the post-course CI results serve as an opportunity for notifying faculty of specific content areas that may need strengthening in future classes. Additionally, faculty instructing subsequent courses will be made aware of potential fundamental knowledge weaknesses of incoming students.

**Development and Validation of New Concept Inventories**

At this time two topics areas in the mechanical engineering core curriculum are not covered by existing concept inventories. It is our intention to commence the initial validation of Machine Design and Manufacturing concept inventories, utilizing similar techniques that have been successfully employed elsewhere to develop previous concept inventories. We plan to pursue the following steps toward development of a valid and reliable instrument:

---

### Test Statistics:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Tests Graded</td>
<td>80</td>
</tr>
<tr>
<td>Number of Graded Items</td>
<td>31</td>
</tr>
<tr>
<td>Total Points Possible</td>
<td>31.00</td>
</tr>
<tr>
<td>Maximum Score</td>
<td>20.00</td>
</tr>
<tr>
<td>Minimum Score</td>
<td>4.00</td>
</tr>
<tr>
<td>Median Score</td>
<td>12.00</td>
</tr>
<tr>
<td>Range of Scores</td>
<td>16.00</td>
</tr>
<tr>
<td>Percentile (25)</td>
<td>9.00</td>
</tr>
<tr>
<td>Percentile (75)</td>
<td>15.00</td>
</tr>
<tr>
<td>Inter Quartile Range</td>
<td>6.00</td>
</tr>
<tr>
<td>Mean Score</td>
<td>11.99</td>
</tr>
<tr>
<td>Variance</td>
<td>13.68</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>3.70</td>
</tr>
<tr>
<td>Confidence Interval (1%)</td>
<td>10.91</td>
</tr>
<tr>
<td>Confidence Interval (5%)</td>
<td>11.17</td>
</tr>
<tr>
<td>Confidence Interval (95%)</td>
<td>12.81</td>
</tr>
<tr>
<td>Confidence Interval (99%)</td>
<td>13.07</td>
</tr>
<tr>
<td>Kuder-Richardson Formula 20</td>
<td>0.63</td>
</tr>
<tr>
<td>Coefficient (Cronbach) Alpha</td>
<td>0.62</td>
</tr>
</tbody>
</table>

### Item Analysis: Answer 1

<table>
<thead>
<tr>
<th>Label</th>
<th>Value</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>39</td>
<td>48.75</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>8</td>
<td>10.00</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>30</td>
<td>37.50</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>3</td>
<td>3.75</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>80</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**Figure 2:** Example Pre-CI Report Displaying Test Statistics and a Single Item/Problem Analysis. The correct answer was ‘B’ for this question.

It should be noted that the post-course CI results serve as an opportunity for notifying faculty of specific content areas that may need strengthening in future classes. Additionally, faculty instructing subsequent courses will be made aware of potential fundamental knowledge weaknesses of incoming students.
1. Identification of foundational concepts for the specified subject area.
2. Determination of student beliefs and attitudes regarding the foundational concepts — validated through student group interview sessions.
3. Formation of response categories — open-ended questions given as quizzes and homework problems in Machine Design and Manufacturing courses to identify misconceptions and incorrect mental models related to the foundational concepts.
4. Formation of multiple choice alpha-test assessment instruments using responses from step 3 as the distracters to test for misconceptions.
6. Determination of non-reliable or weak questions — validated through student group interviews and statistical analysis.

Lessons Learned

Throughout the CI assessment implementation endeavor, the mechanical engineering assessment coordinators encountered several minor tribulations that should be shared with schools who are considering the viability of a similar assessment plan. The following discussion focuses on suggestions that may assist in a smooth transition from current assessment to concept inventory integration.

The first lesson the coordinators learned was to have the concept inventories accessible for faculty review several months before the start of term. This will account for faculty absence, or research conflicts, and allow for adequate time to review the CI exams and prepare the test booklets. Additionally, it is advantageous to have faculty commit to pre- and post-course CI administration dates at the start of the term. As the assessment coordinators discovered with Statics, it is difficult to have faculty relinquish class time for assessment as the end-of-term draws to a close. Furthermore, sending scheduling and exam policy reminders to faculty assessment participants via email is an effective method for ensuring proper assessment administration.

In an effort to assist with the sustainability of the CI assessment program, there are steps that can be taken before and during exam administration. Before administering the CI exams it is essential to determine if the assessment will be scored by University/College testing services, or prepared by the department. By coding bubble sheets with student ID numbers, pre- and post-exam scores can be easily matched and reported for end-of-term grading. It is also helpful to set guidelines for make-up CI exams. It is inevitable that students will add courses to their schedules after the first day of class, and policies should be in place to account for these occurrences.

When administering the CI exam, it is beneficial to preface the exam administration with an explanation of why the CI exams are being given, and the subsequent benefits for students. This provides students with a motivation to perform to their best ability on the assessment. Though students have had several years of experience filling in bubble sheets, our answer sheets were
still plagued with students using an “X”, half-filled bubbles and incorrect answers scribbled out. Therefore, it is recommended that acceptable bubble sheet procedures are emphasized.

Conclusions

Concept Inventory exams continue to gain popularity in science and engineering education because they provide a much-needed rigorous and empirical assessment method for measuring knowledge gains in technical subjects where such metrics are often difficult to develop. Growing out of physics education research, concept inventories have also been developed for many core areas of engineering education. The rigorous procedure for developing the concept inventory exams requires many steps and is quite time-consuming. The ME department at the University of Colorado has gathered eight existing concept inventories that have been developed at universities around the country and matched them to its core curriculum subjects. The exams for five subject areas that appear in the fall 2005 semester of our curriculum were administered during the term with promising results. Three of the five exams were demonstrated to have reliable metrics. One of the five was a beta-exam and we experienced several administrative issues that prevented us from administering the post-test. The other four concept inventories revealed sizable knowledge gains, ranging from 6% to 22% (a 15% to 44% change). We had expected larger gains in the courses where some active learning techniques (hands-on labs, in-class demonstrations or classroom response systems) are utilized, but only Thermodynamics showed the expected strong gain.

Concept Inventories are being integrated into the department’s continuous improvement cycle and have been largely embraced into the culture of the department. This will give the mechanical engineering department a valuable tool that serves many purposes. First, the concept inventories provide critical data for accreditation purposes. Second, the CI reports provide faculty with standard, impartial metrics of knowledge gains, which inform and calibrate both the faculty teaching the core course and also its post requisites. Third, they can drive institutional change by revealing the concepts that students have not mastered and pointing at new methods that may facilitate the teaching of these concepts more readily. Our preliminary experiences and results show great promise for this valuable assessment tool.

Bibliographic Information


