Exploring A New Approach To The Assessment Of Web-Based Materials For Engineering Statics Course

Dr. Paul S. Steif, Carnegie Mellon University

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 education and mechanics. His research has focused on student learning of mechanics concepts and developing new course materials and classroom approaches. Drawing upon methods of cognitive and learning sciences, he has led the development and psychometric validation of the Statics Concept Inventory – a test of statics conceptual knowledge. He is the co-author of Open Learning Initiative (OLI) Engineering Statics, and he is the author of a textbook Mechanics of Materials, published by Pearson.

Dr. Anna Dollár, Miami University

Anna Dollár is a professor in the Department of Mechanical and Manufacturing Engineering at Miami University in Oxford, OH, and previously was on the faculty of the Illinois Institute of Technology (IIT) in Chicago. She received her Ph. D. in applied mechanics from Krakow University of Technology in Poland. Her achievements in engineering education have been recognized by many awards including University Excellence in Teaching Award (IIT), E. Phillips Knox University Teaching Award (Miami University, 2006), and American Society for Engineering Education North Central Section Outstanding Teacher Award (2011). Her research focuses on mechanics of solids and engineering education.
EXPLORING A NEW APPROACH TO THE ASSESSMENT OF WEB-BASED MATERIALS FOR ENGINEERING STATICS COURSE

Introduction

Increasingly it has been appreciated that instruction should be learner-centered. In addition, it is well established that assessment should be integrated into the learning process. In-class assessments, such as peer teaching, minute papers, mudiest-point exercises, and other classroom-based assessments, can give insights into student progress. Computer technology can further integrate assessment into the learning process by offering students individualized, timely help and feedback, which is known to be beneficial. One effort to embed such individualized assessment into learning materials for an entire engineering subject has been the Open Learning Initiative Engineering Statics course.

The OLI Engineering Statics course covers most topics of 2D statics, and consists of 20 modules, akin to chapters, with each module based on a set of carefully articulated learning objectives. The modules contain expository text, diagrams, and simulations, with, most significantly, a large number of interactive exercises. The exercises offer hints and feedback, thereby providing extensive formative assessment to students. These materials were originally conceived of and developed with an independent learner in mind. But these web-based learning materials have also been blended into an instructor-led, lecture-based statics courses using either a flipped classroom format or a traditional format. Student responses on interactive exercises are recorded, which enables the degree to which they have mastered distinct concepts and skills to be tracked. This information on mastery, both for the class as a whole and for individual students, is provided to the instructor in the form of a Learning Dashboard. This allows the instructor to see at a glance which concepts need reinforcement.

Over the course of developing and using the course materials, multiple methods have been employed in an effort to gain insight into their effectiveness. We show and discuss the results of various such efforts summarizing and criticizing/evaluating them in the first part of the paper to provide context for the newly presented work.

Then, we present a new approach to assessing these materials, utilizing a method that has been successfully applied to intelligent tutoring systems. Namely, rather than look at performance over an entire course, we consider successive opportunities to apply various concepts and skills, whether practice improves the likelihood of correct application.

1. Summary of Past Studies of Effectiveness

1.1 Studies of learning gains

Students in a traditional statics courses were assigned to use OLI modules, each containing many interactive exercises. The goal was to isolate the effect of working through the modules setting any influence of lecture or written homework on the same topics. To that end paper and pencil diagnostic quizzes were developed that tested the same concepts and skills as treated by the
module. To quantify learning gains, without the intervening effect of other learning experiences, the diagnostic quizzes were administered in class immediately prior (pre) to completion of the modules and then again in class immediately after (post) completion of the modules. Students received no direct credit for work on the modules, although small credit (with total weight of 2%) was given based on students’ performance on the diagnostic quizzes. Thus, the modules were presented as a tool for learning, rather than to be completed for credit.

Following such a procedure learning gains were studied at two institutions (School 1 Miami University and School 2 Carnegie Mellon University) in different semesters. Five modules had been available when tested at School 1 and nine modules available when tested at School 2. Reported results, reproduced in Table 1, included pre- and post-scores, gains and normalized gains defined as $G = \frac{(Post - Pre)}{(Max - Pre)}$ (corresponding to the actual increase in score compared to the maximum possible increase).

Students at the two Schools have different preparation and prior experience with Statics, explaining the different pre test scores. Nevertheless, high normalized gains, varying from 0.45 to 0.80 were found at both institutions, suggesting that the materials are appropriate for students with various levels of preparation. Hake used $G$ to compare scores on the Force Concept Inventory from different institutions; gains of about 0.5 were relatively high, typical of classes with more interactive engagement.

### Table 1. Pre-post gains in diagnostic quiz scores due only to use of OLI modules.

<table>
<thead>
<tr>
<th>Module</th>
<th>Miami University</th>
<th>CMU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post</td>
</tr>
<tr>
<td>1</td>
<td>38%</td>
<td>81%</td>
</tr>
<tr>
<td>2</td>
<td>51%</td>
<td>94%</td>
</tr>
<tr>
<td>3</td>
<td>38%</td>
<td>70%</td>
</tr>
<tr>
<td>4</td>
<td>45%</td>
<td>66%</td>
</tr>
<tr>
<td>5</td>
<td>21%</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

High, statistically significant normalized learning gains based on diagnostic quizzes were found, varying from 0.4 to 0.78 over the 9 modules tested, for students of both low and high pre-test scores.

1.2 Studies of correlation between OLI quizzes’ scores and exams & statics concept inventory

The diagnostic quizzes used for studies of learning gains were later embedded into the OLI course as end-of-module assessments. Furthermore, these quizzes became the only activities that
were formally reported to a gradebook and could thereby be used to assign a grade. A practical way of assigning use of the OLI course, as done in multiple classrooms, is then to instruct students to complete as many activities within each module as they saw fit to learn the material and perform well on the quiz. Yet, it remained to be seen whether the quizzes were good indicators of whether students had learned the material as needed for performance in exams.

Using data from classrooms at three institutions, Pearson correlations were computed between the mean OLI quiz scores and two summative of learning in the course: exam scores and the Statics Concept Inventory (SCI) administered towards the end of each course. As seen from results reproduced in Table 2, quiz scores are highly correlated with other, more valued, course performances such as exams and SCI scores. By contrast, scores on written homework had very low correlations with exams and SCI scores. Quiz scores could, thus, even be used as early warning signal for course performance.

Table 2. Pearson Correlations at three institutions between mean OLI quiz score and performance in course (exams at Miami and CMU, final grade at Itasca) and on Static Concept Inventory.

<table>
<thead>
<tr>
<th></th>
<th>OLI Quizzes- Course Performance</th>
<th>OLI Quizzes-SCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMU</td>
<td>0.61</td>
<td>0.49</td>
</tr>
<tr>
<td>Itasca</td>
<td>0.80</td>
<td>0.74</td>
</tr>
<tr>
<td>Miami</td>
<td>0.45</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Scores on online end–of–module quizzes have been positively correlated with performance in class exams and with scores on Statics Concept Inventory. The evidence suggests that quiz scores in OLI Statics might be used as an early warning signal regarding final course performance and conceptual understanding.

1.3 Studies of correlation between course performance and usage of available exercises

Given the wide variation in student performance on the OLI quizzes, on the SCI, and in the course exams, we have sought to determine whether this performance can be explained by usage of the available activities within a module.

Because of OLI’s original design as a tool for individual learners, there are far more activities than most students would normally undertake. The students in the classrooms where instructed to use only as much as they deemed necessary. Despite their not being required, in-module exercises were nevertheless undertaken by many students: typically from 50% to 75% of activities were initiated by students, with standard deviations on order of 20%. This implementation in which students chose which activities to undertake with the goal of learning the materials and scoring well on the quiz, gives students significant responsibility for self-regulating their learning.

Moreover, the factors that affect how much students use optional materials are many and these push students towards more or less usage independent of their perception of whether they have learned enough. Accordingly, we studied via survey the factors that affected students’ optional engagement in activities. In general, we found in that students were attracted by the frequent feedback and its benefits, including the explanations that came with wrong answers, the
availability of hints that scaffolded their learning, and the way the resource succeeded in breaking down the material into distinct concepts. Students primarily objected to the large amount of reading associated with using the materials (readings that were devised to weave together the ideas for an independent learner). Few students objected to being frequently assessed.

Two studies \textsuperscript{11,12} considered the percentage of the activities undertaken within a module, and whether that percentage explained the variation in performance. In practice, quantifying the engagement of students in the materials was limited to knowing the fraction of activities in each module (chapter) initiated by each student. Moreover, initial ability of students was also considered by using a pre-course Force Concept Inventory score. These studies revealed that the fraction of exercises initiated was itself not strongly related to class performance. In short, we could not consistently find statistically significant relations between the fraction of activities initiated and most measures of performance. Perhaps the hypothesis that performance would correlate with usage was too simplistic: learning may depend on how an individual student processes the feedback that accompanies the exercises. Moreover, if students were truly self-regulating as requested, using the materials only as needed to learn, then an increase in performance with usage should not necessarily be expected.

\textit{In response to assignments to undertake in-module activities only as needed to learn and perform well on the quizzes, a wide variation in usage was found, but the number of activities undertaken was not consistently correlated with performance.}

2. Learning Curves Analysis to Gauge Relation between Usage and Learning

Earlier efforts just described, which failed to detect a specific effect of engaging in exercises, may have taken too large a view: namely looking the influence on overall performance in the course. Here we present an alternative approach to evaluating the benefits of engaging in the OLI interactive exercises based on Learning Curve analysis, a technique used for intelligent tutoring system. Most of the exercises can be viewed as composed of a set of questions each of which the student answers correctly or incorrectly. We need to posit, for each such question, one or more pieces of knowledge (referred to as knowledge components) that one must possess to answer the question correctly. We are interested in evidence that knowledge components are being acquired (learned) as a consequence of doing the exercises and receiving feedback. This is simplest to explain in the case where we have a set of questions that can be answered correctly with only the same single knowledge component. We track each student’s efforts to answer those questions, that is, efforts to exercise that knowledge component. Then, for a group of students, we look at the percentage of that make an error in the first attempt to use the knowledge component, the second attempt, and so forth. We then plot the percentage in error as a function of the attempt. Note that an attempt to use a knowledge component could correspond to different questions, depending on the student. Still, \textit{if the same piece of knowledge is being tapped, and if there is value in getting feedback on each attempt}, then we would expect the percentage in error to decrease for successive attempts.

The learning curves, or plots of the percentage in error as a function of attempt, are often quite noisy. Thus, besides the plots, we use the tools available in the Pittsburgh Science of Learning
Center data repository Datashop to fit a regression model commonly used for learning curves, in which the probability of making an error (logistic of the probability) is predicted as a linear function of the attempt. For a single knowledge component, the statistical model takes the form:

$$\ln\left(\frac{1 - e_{ij}}{e_{ij}}\right) = \theta_i + a_j + b_j \cdot T_j$$

Where $e_{ij}$ is the probability that student $i$ makes an error in applying the jth knowledge component, and $T_j = 1, 2, 3, \text{etc.}$ corresponds to the first, second, third, attempt to apply the jth knowledge component. The parameters in the statistical model correspond to a single parameter for each student ($\theta_i$), an initial ease in applying each knowledge component ($a_i$), and a slope or decrease in the probability of an error with practice for each knowledge component ($b_i$). The full version of this statistical model implemented in Datashop allows one to associate multiple knowledge components with a question.

We sought to adopt this learning curve approach: with OLI Engineering Statics we have software that offers students opportunities to solve problems (answer questions) and to receive hints and feedback along the way. (Note that many activities are not in the form of questions per se, but steps to solve problems, such as clicking on a force direction, entering magnitude, choosing from different rationales, and so forth. Each of those steps where there is a correct response is considered a “question” for this discussion.) To quantify the effectiveness of the learning activities we had to hypothesize a knowledge component model of the domain (the set of knowledge components needed to answer questions) and to associate each question with specific knowledge components. Then, the fit of the logistic regression enables one to determine empirically whether the knowledge component model is supported by the data and whether usage of the software is helpful, in the sense that students are gradually making fewer errors in successive attempts to apply what are believed to be those same skills.

From all the activities in the OLI Engineering Statics course there are more than 1100 questions (opportunities for students to respond), spanning 20 modules or chapters, which are recorded. Many of the course activities were devised prior to the prospect of using a learning curve analysis arose. The first knowledge component model consisted of slightly over 100 knowledge components; approximately 200 questions posited multiple knowledge components. The skills are not always fundamental concepts, but operational knowledge needed to respond to questions as they were posed. By design the initial OLI activities in many modules would address the basic skills needed to solve problems – a single concept or skill would be needed. Examples of such single skills are: “identify point on a body at which a force acts”, “represent the interaction due to an attached cord”, or “determine a moment arm”. Sometimes students would be taken through the steps to solve problems, usually needing one skill for each step. Finally, activities would involve solving full problems, where a number of skills were needed, but students only entered the final answer. In such cases, students would need many skills – those for all steps. In such instances, we chose to associate with the question (the final answer) a single composite skill, for example “solve a 2-D statics problem” or “Compute load to cause tipping”, where in both cases the student needed to do all steps of drawing a free body diagram, writing down equations, solving and interpreting them, to provide a single answer.
Several learning curves are shown in Figures 1-4, for a cohort of 74 students who used the OLI Engineering Statics as part of a lecture based statics course. The blue curve is the prediction based on the model fit. These illustrate that for some knowledge components one observes a statistically discernable decreasing error rate (Figures 1 and 2), while not for others (Figures 3 and 4). Opportunity 1 typically involves all students. Then, because students do different numbers of problems in different orders, the number of students represented in subsequent opportunities drops off. Typically the very wild swings at high opportunity number (Figure 3), correspond to one or two students exercising that skill.

Figure 1. Learning curve for skill of relating the sign of a moment to the sense of rotation produced by a force.

Figure 2. Learning curve for skill of recognizing that a couple involves zero net force.
In OLI Engineering Statics we found decreasing error rates in some of the skills, but not for the majority of skills. We attribute this finding to three factors. First, the responses to some questions are posited to exercise just a single skill; however, with others, the response is legitimately viewed as exercising multiple skills. This is consistent with our wanting students to integrate their knowledge to respond to some problems. So, the same simple skill might be exercised in isolation and then as part of a question that requires other skills as well. While the statistical modeling associated with Learning Curve Analysis allows for multiple skills to be reflected in the response to a question, the desirable decreasing error rate seems to be particularly rare in the case of skills that are sometimes exercised in conjunction with other skills. Second, because the materials cover an entire semester-long course, it is difficult in many cases to have students...
undertake the multiple attempts at each skill that are likely needed to see statistically significant decreases in error rates. Third, the difficulty in resolving the learning of each skill is further exacerbated by the fact that different situations of applying what we deem to be the same skill may be more or less difficult. Of course, one could try to further differentiate between skills, but as it is there are approximately one hundred across the course as a whole. And, then it will be more likely that there are insufficient opportunities to practice the different skills.

3. Summary and conclusions

The OLI Engineering Statics course offers web-based instruction that provides many embedded interactive exercises offering individualized feedback and help to students, while giving instructors summary information on student performance. Various efforts to assess the effectiveness of courseware have been presented in the past and summarized here.

As demonstrated in past research, students exhibit significant gains in diagnostic quizzes after using OLI modules. Also, quiz scores have been shown to correlate positively with other important measures of performance, such as exams and statics concept inventory scores. However, it has been found that overall course performances do no correlate with how many optional exercises a student initiates. This result was to some extent expected, since in these studies students were not required to undertake a certain number of exercises, but rather to do as many as they deemed necessary to perform well on the end-of-module quiz. Student usage was indeed found to vary widely, indicating self-regulation, which lessens the likelihood of a positive correlation between the number of exercises undertaken and performance.

As an alternative to overall performance in a course, one can look at a much smaller grain size and potentially detect whether repeated exercises on specific concepts and skills do lead to fewer errors. With the goal assessing the benefit of engaging in specific exercises, a new approach has been presented here, which adapts the method of Learning Curve Analysis that is common in intelligent tutoring systems. Each individual question within each exercise is tagged as corresponding to one or more skills. Then, successive opportunities for students to exercise each skill are tracked. If feedback on each opportunity is effective, then the students should make fewer errors in successive opportunities to practice a given skill.

It has been found that for some skills there is clearly improvement with practice. For other skills, the variation with practice is much less clear. Various reasons for the absence of improvement were proposed. Interactive exercises often require multiple skills, which are harder to disentangle, sometimes there might be insufficient opportunities to practice a skill, and sometimes different opportunities involve applications of the same skill that are arguably quite distinct in difficulty.

Methods for gauging the effectiveness of web-based materials should continue to be sought, including comparisons with alternative instruction. Yet, when overall performance is taken to be the basis for comparison it is difficult to pinpoint where specifically the materials could be improved. Learning Curve Analysis offers a potentially valuable means of determining whether students are making progress in acquiring specific skills addressed by a collection of exercises, and thereby point to particular exercises where improvements should be sought.
ACKNOWLEDGEMENTS

The authors appreciate the support provided by the National Science Foundation through grant DUE 0918271, the William and Flora Hewlett Foundation through the Open Learning Initiative at Carnegie Mellon University, by the Department of Mechanical Engineering at Carnegie Mellon University, and by the Mechanical and Manufacturing Engineering Department at Miami University.

Bibliographic Information

6. A. Dollár, P. S. Steif, Web-Based Statics Course Used In An Inverted Classroom; Proceedings of the American Society for Engineering Education Annual Conference & Exposition, Austin, Texas, June 2009