Study of Usage Patterns and Learning Gains in a Web-based Interactive Static Course

PAUL S. STEIF
Carnegie Mellon University

ANNA DOLLÁR
Miami University

BACKGROUND
Courseware for engineering education can feature many discrete interactive learning elements, and typically student usage is not compelled. To take advantage of such courseware, self-regulation of learning may be necessary. Evaluation of courseware should consider actual usage, learning gains, and indications of learning self-regulation.

PURPOSE (HYPOTHESIS)
The research question focuses on how students' interactions with the courseware affect their learning gains. The hypothesis tested is that learning gains from online courseware increase with usage, and particularly with usage that suggests learning self-regulation.

DESIGN/METHOD
Students in a lecture-based statics course were assigned to study previously developed courseware as part of homework assignments. Learning gains were deduced from pre- and post- paper and pencil diagnostic quizzes, and from the first class exam. Credit was based on quiz scores, rather than courseware usage. Usage of interactive elements of the courseware was inferred from log files of students' interactions with the courseware, and patterns suggesting learning self-regulation were identified.

RESULTS
High, statistically significant learning gains were found. Substantial usage was evident, with core learning activities initiated by, on average, three-quarters of students. Learning gains and performance on the relevant class exam appeared to be more closely correlated with usage that indicated self-regulation of learning rather than with total usage of the courseware.

CONCLUSIONS
Methods of assessing courseware should go beyond courseware features, learning gains, and student self-reports of effectiveness to include monitoring of actual usage and analyses relating usage to learning. Self-regulation of learning is likely to be critical to successful usage of courseware, and courseware should be designed to encourage it.

KEYWORDS
courseware, online learning, statics

I. INTRODUCTION

A. Background and Rationale

Computer-based learning materials play an increasing role in the education of engineers. Courseware of various types has been and will be used in educational contexts. These include Learning Content Management Systems, such as Blackboard, which can make a variety of resources, often reading materials, available to students. There are also simulation environments, which allow students to virtually explore aspects of different domains. These might include realistic systems such as civil engineering structures (Martini, 2006) or idealistic systems such as electric charges in a plane (Sherwood and Chabay, 1991). Courseware also includes problem solving environments, for example the Andes tutor for physics problem solving (Van Lehn et al., 2005). Although they have not yet been developed for engineering subjects, cognitive tutors, which address subjects such as high school algebra and geometry, are problem solving environments that are based on a cognitive model of expert and student problem solving. All actions of students while working with the cognitive tutor are tracked and used to alter the future course of instruction.

This paper focuses on courseware that occupies a middle ground among the types described. It seeks to enact instruction by providing carefully designed sequences of short text, graphics, and interactive activities. The interactive exercises include simulations that help to elucidate physical phenomena, activities that allow practice of problem solving procedures with help and feedback, and tests of students' comprehension.

To ensure that computer-based learning materials contribute significantly to learning, we should assess their effectiveness, and, based on those assessments, employ methods for improving the materials. Effectiveness of courseware has been assessed in various ways. Two common ways include: (i) obtaining student self-reports of their perceptions of effectiveness, and (ii) utilizing an independent diagnostic measure, which is administered before and after the period of use (pre- and post-tests) and then measuring gains. While students can legitimately comment on whether they enjoyed using the courseware, self-reports of learning effectiveness are notoriously suspect (Salaberry, 2001). While the pre-post gain is useful, it alone may miss information on the extent to which the materials were actually used. For example, in instances where the courseware is obviously used to complete an assignment it generates output that is specifically needed for the assignment, then use is known, but in other cases the amount of use is largely unknown.

Assessment of engineering courseware should recognize two factors peculiar to its use. First, since problem solving occupies a central role in many engineering courses, successful courseware will likely contain interactive elements that seek to promote the acquisition of problem solving skills. Second, consistent with the practice in higher education, usage of educational resources is usually not compelled or is at most indirectly compelled, so the precise actions taken by students using educational materials do not occur under the eye of the instructor. Further, where learning materials intersperse text with activities that require student interaction and response, there is evidence that merely reading the text and bypassing the activities leads to markedly less learning (Arnold et al., 2005). Thus, page views by themselves, without monitoring engagement in activities, will be an insufficient measure of usage. Furthermore, it is perhaps wise to heed the warning (Fisher, 2007) that availability of learning opportunities is not to be confused with actual use. In summary, for learning to occur with engineering courseware it is critical that students use interactive instructional elements, although given the nature of assignments in engineering courses, such usage is not guaranteed.

Tracking the usage of computer-based learning materials in particular is feasible, since all interactions can in principle be preserved electronically. The preserved data offer unprecedented access into the learning activities of students, and the subsequent analysis of the preserved data is becoming a field in its own right, educational data mining, with a first International Conference held recently (Educational Data Mining, 2008). Educational data mining seeks to analyze large datasets to answer questions in educational research. Such large datasets might emerge from intelligent tutoring systems or Web-based educational courseware. The methods of educational data mining can be applied to obtain results that are of interest to students, to instructors, or to administrators (Romero and Ventura, 2007). Certain types of analyses are common in educational data mining: cluster analysis (identifying objects that should be grouped), classification (developing procedures for classifying an object into one of several groups), and association rules (determining whether items that appear at the same time are meaningfully related). Usage statistics are a common starting point in many educational data mining studies.

Usage statistics include simple measures such as the total number of visits or number of visits per page (Pahl and Donnellan, 2003), or the participation of learners in multiple learning sessions over time (Zorrilla et al., 2005). Quantification of usage might focus on time spent solving individual problems (Nilakant and Mitrovic, 2005; Warnakulasooriya, Palazzo, and Pritchard, 2007). In Computer aided Language Learning, tracking data have been used for various purposes, including completion of assignments (Opp-Beckman and Kieffer, 2004), assessment of learning (Goodfellow, 1999), identification of problem areas for follow-up learning activities (Colpaert, 2004), determination of the optimum number of participants in computer-mediated communication (CMC) sessions (Böhlke, 2003), examination of students' degree of participation in CMC sessions (Chun, 1994), and how students interact with specific software components (Chapelle, 2001, 2003). Many more examples could be cited, but it is important to emphasize that comparable studies of the usage of engineering courseware do not appear to exist.

In instances where usage is not compelled and significant numbers of activities are available, and (as is usually the case) the needs for instructional support differ across students, effective usage requires students to self-regulate their learning. Self-regulation includes meta-cognition (thinking about one's thinking), strategic action (planning, monitoring, and evaluating one's progress against a standard), and motivation to learn (Winne and Perry, 2000). In courseware that offers a significant number of varying instructional elements in particular, students need to choose which among the available activities to pursue based on their own assessment of their learning. Meta-cognitive abilities are consistently cited as necessary to the repertoire of successful learners (Brown, 1980; Flavell, 1991). Students have different native meta-cognitive abilities, and there have been efforts to further develop these abilities in particular domains through specific instructional programs (Campione, Brown, and McConnell et al., 1988; Scardemalia, Bereiter, and Steinbach, 1984). Moreover, one can seek to design courseware to encourage some level of self-regulation or provide interactions that scaffold self-regulation. In any event, when self-regulation is believed to be important to success in using courseware, assessment of courseware may seek evidence of self-regulation and how its presence is reflected in learning gains or performance. The analysis of usage patterns, in particular, represents one potentially fruitful avenue for obtaining indirect evidence of self-regulation.

B. Proposed Contribution and Research Question

This paper addresses a relatively new class of engineering courseware which contains many discrete interactive instructional elements that are interspersed into text; learning likely depends critically on using the interactive elements. Moreover, in many, if not most instances, students using these materials in a course will choose whether or not to use individual elements, rather than be compelled to do so. Thus, this paper adds to the literature of courseware assessment by focusing on new complex courseware, which is utilized by students who control the specifics of their usage. Moreover, the paper joins the concept of self-regulation, known to be relevant in other learning contexts, to courseware assessment. In particular, we hypothesize that self-regulation by students of their usage of such materials is related to learning gains.

The research question that we seek to answer is whether learning gains are related either to overall usage or to self-regulated usage of the online learning materials. To answer this question, we need to analyze usage statistics, which requires a high level of monitoring and subsequent analysis that, while common in other fields, has yet to be conducted in the context of engineering courseware to our knowledge. The contribution of this paper is to propose and extract relevant measures of usage for new complex courseware, to identify markers of self-regulation, and to determine whether learning gains are tied to overall usage, or self-regulation in usage.

C. Overview of Presentation of the Paper

In Section II, we will describe (A) the participants in the study, (B) the data collection methodology, (C) the Web-based courseware, "Open Learning Initiative" (OLI) Engineering Statics Course, used in this study, and (D) how the data are to be analyzed to address the research question. In Section III we will present the results, which will be followed by a summary in Section IV.

II. Method

A. Participants

The participants in this study were 110 college students who were enrolled in a lecture-based, semester-long statics course in the

Department of Mechanical Engineering at Carnegie Mellon University in Fall 2007. As part of regular homework assignments during the first six weeks of the course, students were assigned to complete online one to two modules per week of the OLI Engineering Statics course, with nine modules assigned in total. No credit was given for actual usage of the OLI course materials. Rather, pencil and paper diagnostic quizzes were devised to address the concepts covered by the modules and were administered in class before (pre) and immediately after (post) completion of each week's modules. No lectures or other homework covered the same material prior to the pre- and post-quizzes. Students were informed about the diagnostic quizzes and that they would count towards their course grade (with total weight of 2 percent). Thus, the modules were presented as a tool for learning, rather than to be completed for credit.

B. Data Collection

The servers on which OLI courses are running log the majority of students' interactions (reaching a page, initiating an activity, choosing answers or invoking hints), and the resulting log files are available for subsequent analysis. In the future, it is anticipated that this Web-based course, and likely many others, will have an instructor digital dashboard, where student participation in various learning activities can be monitored in real time. Because of the massive amount of data (100 students, 300 learning activities, some of which may necessitate many student interactions), complete analysis is difficult; instead we quantified specific aspects of usage that serve to address the research question posed earlier. Accordingly, log files corresponding to modules 1–5 have been studied to various extents.

Two performance measures were used, the pre-post diagnostic quizzes, and the scores from the first class exam. The grade in the course is based primarily on the four class exams, and nearly all the material covered by modules 1 to 5 were included in the first exam (along with additional material). Nearly all students in the class adopted unambiguously identifiable user names for the OLI course, and so performance in the course (diagnostic quizzes and exams) could be related to usage of the OLI course materials.

C. Description of the Courseware Used in the Study—Open Learning Initiative Engineering Statics Course

CMU's Open Learning Initiative (OLI), supported by the William and Flora Hewlett Foundation, has sought to develop cognitively informed, Web-based courses for introductory, college-level subjects. OLI courses are freely available to anyone through the OLI Web site (Open Learning Initiative, 2009). In addition, an instructor can register a class, so that the students' activities on the OLI course are visible to the instructor. Engineering Statics is the only engineering course in the suite of OLI courses.

Because of its importance in engineering curricula, statics continues to enjoy broad attention from the engineering education community. Furthermore, learning in statics is viewed as inadequate: instructors in design courses bemoan the difficulties students have in applying statics as a practical tool in their design projects (Harris and Jacobs, 1995), and performance by students from a variety of institutions in the Statics Concept Inventory (Steif and Hansen, 2006, 2007) indicates that many students have an inadequate grasp of essential conceptual components of statics. As explained in detail in Dollár and Steif (2008), the OLI Engineering Statics course rests on prior efforts to monitor and assess student learning difficulties in this subject (Steif, 2004, Steif and Hansen, 2007), efforts to reinvent the teaching of the subject (Steif and Dollár, 2005), and classroom materials developed to address those difficulties (Dollár and Steif, 2006). As catalogued in more detail in Dollár and Steif (2008), the instructional approaches pursued in the OLI Engineering Statics course were informed by general lessons from the learning sciences, including the benefits of feedback during learning and scaffolding which can fade as mastery occurs.

The OLI Engineering Statics course currently consists of a series of five units, which comprise sixteen modules. A module consists of a series of pages, each devoted to a carefully articulated *learning objective* that is independently assessable.

Within each page, relevant concepts, skills, and methods are explained using not only words and static images, which are typical of textbooks, but also through additional means that engage learners in active learning. Since an ultimate goal of the course is to apply statics to genuine artifacts, a competence necessary to engineering practice, the course seeks to take advantage of digital images of relevant artifacts and video clips of mechanisms. Consistent with the authors' pedagogical philosophy of focusing initially on forces associated with manipulating simple objects, students are often guided to manipulate simple objects to uncover relevant lessons. The interactive elements of the course (about 300 total) include simulations, walkthroughs, and tutors with hints and feedback, which are now described.

Simulations, which generally involve motion, are used extensively to convey basic concepts in statics, consistent with the authors' pedagogical philosophy of making forces and their effects visible. Some simulations are *non-interactive*, and are simply played by the student; other, *interactive, guided simulations* allow students to adjust parameters and to see their effects (what-if analysis). These are often initiated by a question posed to the student. Simulations help learners connect calculations and numbers with physical representations. After each simulation, there is a short "Observation" to ensure that the student takes away the intended lesson of the simulation. Figure 1 shows an example of interactive guided simulation from module 5.

Since statics is a subject that requires solving problems as well as understanding concepts, larger tasks have been carefully dissected and addressed as *individual procedural steps*. Often, the application of the procedure is demonstrated with a "Walkthrough:" an animation combining voice and graphics that walks the student through an example of the procedure (see Figure 2). Such an approach is viewed as particularly effective, since it engages both aural (hearing) and visual pathways, diminishing the mental load on each. This is particularly the case when we want the student to make appropriate connections between words and evolving graphics.

Students themselves engage in problem solving procedures first in formative assessment Learn By Doing (LBD) exercises. These are computer-tutors in which students can practice a skill as they receive detailed, individualized, and timely hints and feedback. Usually, the student can request a hint at each step. If there are multiple hints, often the first hint reminds the student of the relevant underlying idea or principle, the second hint links the general idea to the details of the problem at hand, and the final hint virtually gives the answer away, but explains how one would arrive at the answer. Wrong answers at each phase provoke feedback, which may be generic ("That's not right") or tailored to each incorrect answer, particularly when a likely diagnosis of the error can be made. Some computer-tutors offer scaffolding: the user can work independently towards the solution or request help, consisting of a series of substeps. The user can go back and try to answer the main question at

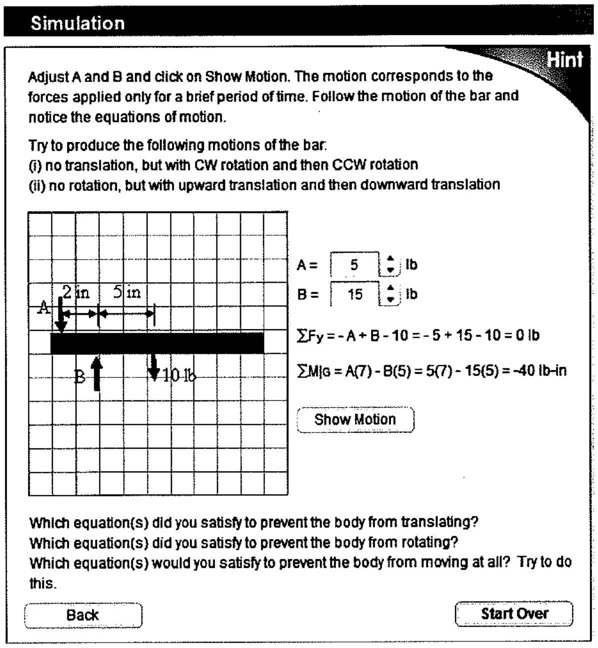

Figure 1. Guided simulation motivating the discovery that forces on a body independently control the translational and rotational tendencies, and that both must be zero for equilibrium.

any time. All activities can be engaged several times by students, and in some instances, multiple versions of a problem are generated with new parameters to enable further practice.

At the end of each page, students can test if they have grasped concepts and mastered procedures through tutors that are referred to as "*Did I Get This?*" (DIGT) exercises. These tutors are similar in form to LBD tutors and offer students a last chance (a summative assessment) to determine if learning objectives were met, or whether further study of previous material is warranted. Figure 3 shows two screenshots from a DIGT tutor from module 4, which asks the student to calculate the moment of a force using a component perpendicular to a given distance, and offers scaffolding.

Thus, a typical page would first state the learning objective, followed by some introductory text. Then, there might be a simulation together with some text to explain a concept. Next, a procedure might be presented and then demonstrated with a "Walkthrough". Students then practice the procedure, or test their conceptual understanding with one or more LBD tutors. Finally, there are one or more DIGT tutors, which allow students to check their understanding, followed by a brief synopsis of the key points, referred to as *To Sum Up*. Figure 4 shows an example of a typical OLI Engineering Statics page, with the navigation bar on the left. The screenshot captures the lower half of a particular LBD with a hint pertaining to the current step, a link to a summative assessment exercise (DIGT), followed by a To sum up. Towards the bottom is the "My Response" link from which students can give feedback on the page, and the link from which the instructor can access the student feedback.

D. Approach to Data Analysis

The central research question focuses on how learning activities in courseware affect learning gains, including activities indicative of self-regulation. To pursue this question, we first establish the presence of significant learning gains, as measured by the diagnostic quizzes. Second, through an analysis of log files, we quantify usage of the interactive instructional elements of the courseware, seeking to identify usage patterns that suggest self-regulatory processes. Third, we analyze performance data, namely diagnostic quizzes and exam one scores, in conjunction with usage data, and seek to identify whether usage in general and/or usage suggestive of self-regulation are tied to performance.

Learning gains for each of the modules were established based on pre- and post-test scores from diagnostic quizzes for which normalized gains were computed. Usage statistics, and in particular the variation in usage across students, focuses on the frequencies at which interactive elements of the course are initiated and completed. Thus, rather than page views, which are commonly quantified in studying Web-based materials, the analysis here focused on those components of the course that go substantially beyond traditional text and graphics, such as simulations, Walkthroughs, and tutors with hints and feedback. Participation in learning activities can be defined in many ways: the initiation of an activity, any clicks or text entries while working on the activity, and the completion of the activity. The course offers significant numbers of learning activities, and since instructional needs differ among students, effective usage requires students to self-regulate their learning. A student who worked through several learningactivities may choose to engage in additional ones, or may decide not to do any more, based on the student's gauging whether sufficient learning has occurred via the activities completed so far. Self-regulation means that students are choosing which among the available activities to pursue based on their own assessment of their learning. We call particular attention to usage patterns that appear to signal self-regulatory learning processes.

In the first five modules of the course, which were available at the time of the data collection, there are 106 interactive elements: 75 tutors with hints, feedback and scaffolding (34 LBDs, 41 DIGTs), 9 "Walkthroughs" and 22 simulations. This represents a massive amount of data; time and resources have thus far permitted the analysis of only a subset of these data. Currently, the system logs all student interactions with pages and with individual interactive elements (with each click and text entry being logged). At the time of this work, the logged data resided in a database on which specific queries could be run; the output of such queries were analyzed to produce the results presented here.

The relationships between usage of the interactive elements of the course and learning gains are studied as follows: since only module 5 had a high standard deviation in post-test scores on diagnostic quizzes, efforts to relate usage to gains focus on module 5. In particular we consider module 5 diagnostic quiz gains vs. overall tutor use in module 5, as well as performance on individual questions in the diagnostic quiz vs. overall tutor use and the use of specific potentially relevant tutors. Then, since patterns of use of LBD and DIGT tutors are so distinct, we consider how gains correlate with LBD and DIGT use separately. Finally, since class exam 1 included all the material covered by modules 1 through 5, the separate correlations between LBD and DIGT and exam 1 performances are studied. Again, particular attention was focused on assessing whether overall usage or usage consistent with self-regulation usage is more correlated with performance.

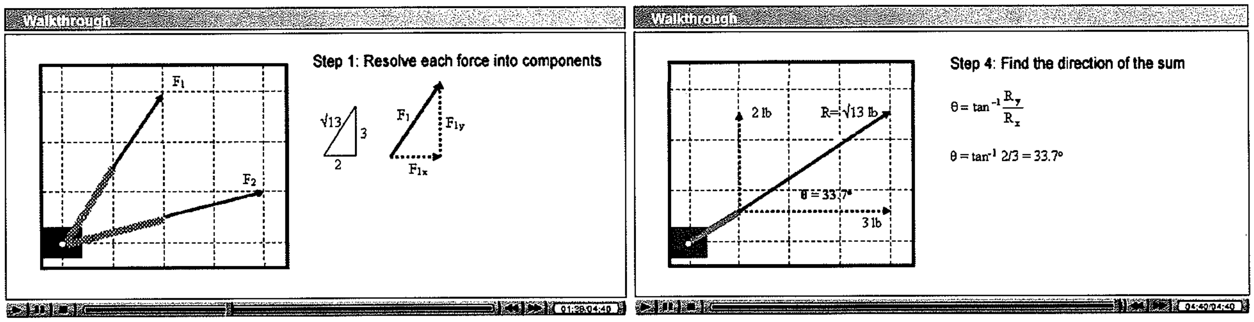

Figure 2. Selected screenshots of a "Walkthrough" from module 4.

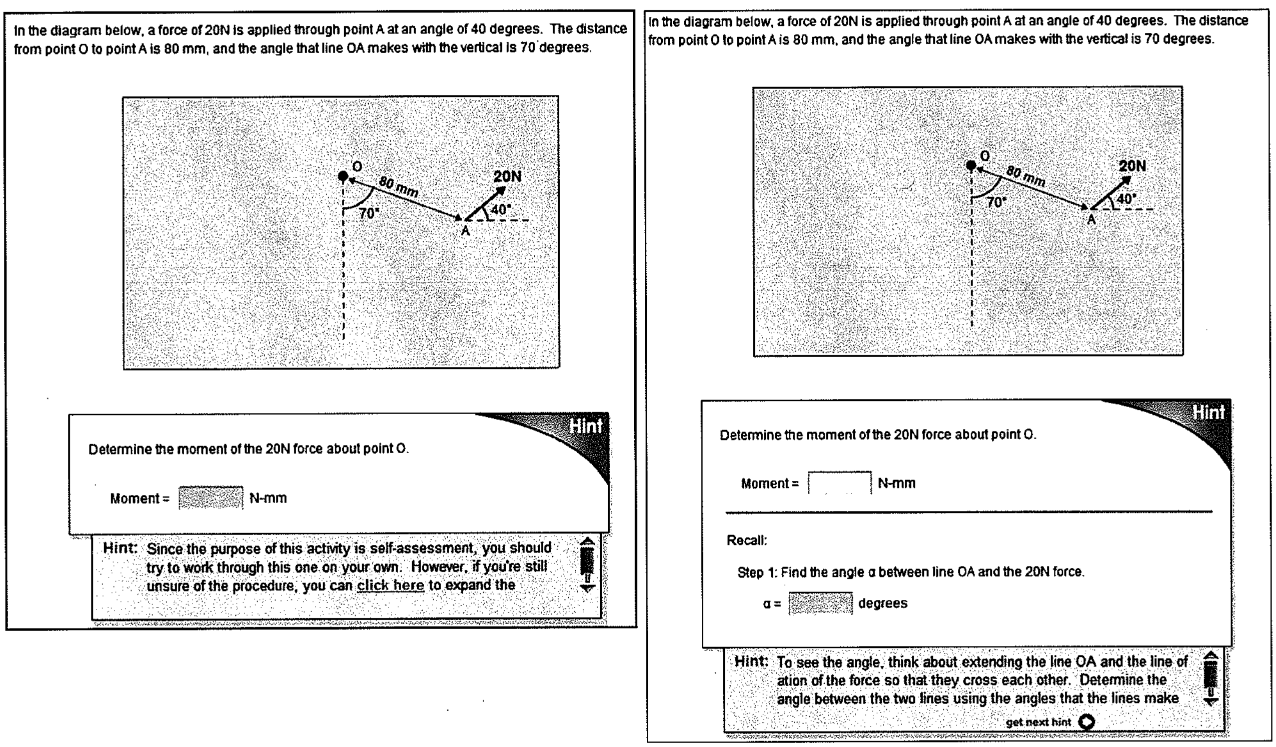

Figure 3. DIGT with hints, feedback, and scaffolding from module 4.

III. Results

A. Learning Gains

Results are presented here for the 103 students (out of 110) in the Carnegie Mellon University class who took both the pre- and post-assessment quizzes given in class. As measured by the paper-and-pencil assessment tests, the learning gains pre to post were significant. Mean pre- and post-test scores, gains, as well as normalized gains, G, which varied from 0.40 to 0.78, were computed and are displayed in Table 1. The normalized gain is defined as

$$G = \frac{(\text{Post} - \text{Pre})}{(\text{Max} - \text{Pre})} \tag{1}$$

Normalized gain, used for example by Hake (1998) to compare data from the Force Concept Inventory (Hestenes, Wells, and Swackhamer, 1992) from different institutions, corresponds to the actual increase in score compared to the maximum possible increase. Gains of about 0.5 in Hake's study were considered very high. The gains reported for Miami University (Dollár and Steif, 2008), were comparable to those reported here, even though the pre-test scores were lower (CMU students have a three-week introduction to statics during a freshman mechanical engineering class). This suggests that the materials are appropriate for students with various levels of preparation.

As a further test of the statistical significance of the normalized gains, the standard error of the mean (SEM) was also calculated (see Table 1). This is an estimate of the error in using the sample to estimate the mean normalized gain of the population. The SEM is much less than the gain. When the mean of the normalized gain for each module is compared with zero via a t-test, the probability that normalized gains do not differ from zero is less than 0.0001. These both suggest that gains are highly significant statistically.

Because the mean post-test scores for module 5 were lower than the preceding 4 modules, this module was singled out for

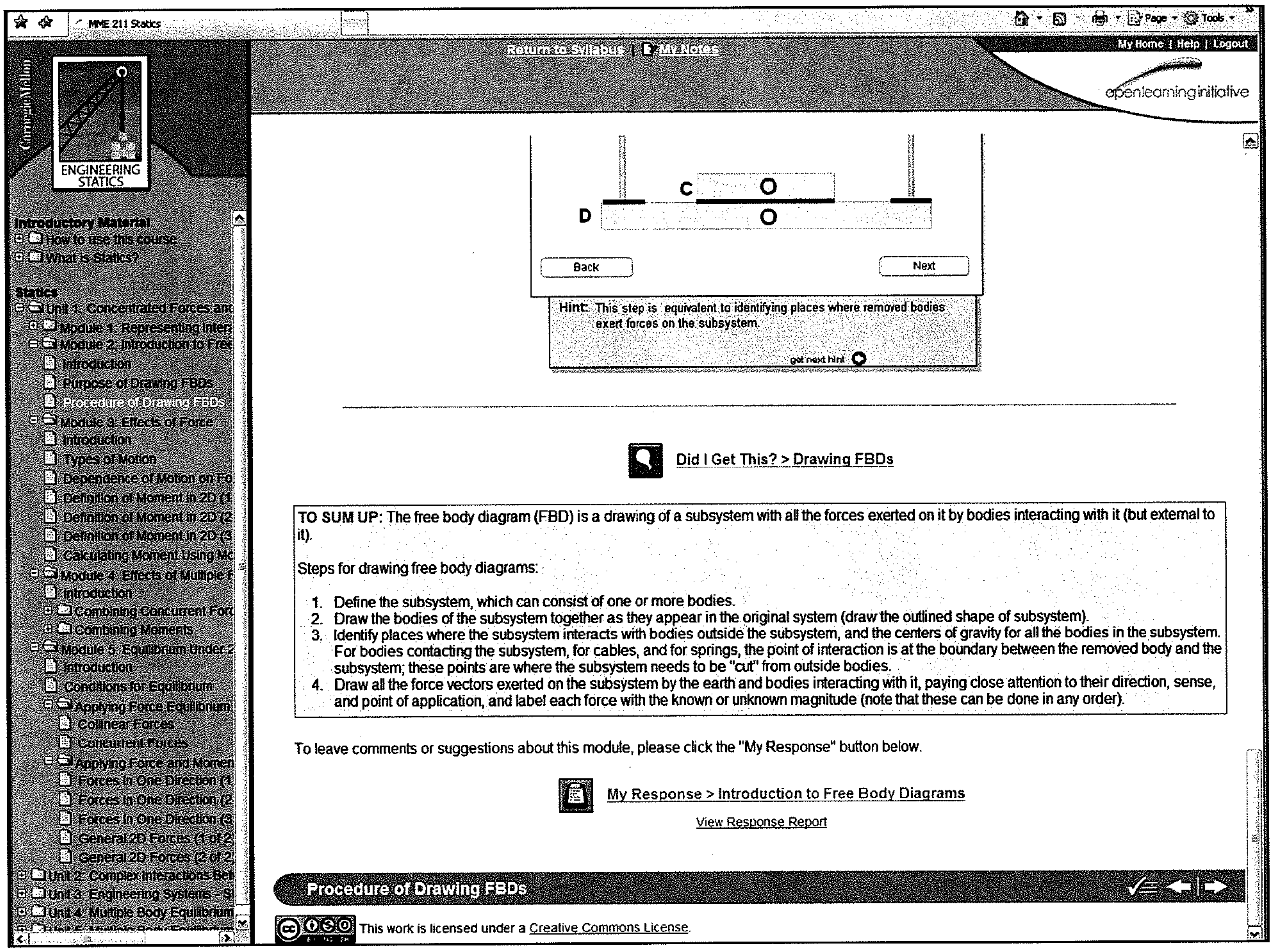

Figure 4. Typical page from the OLI Engineering Statics course.

Module	Pre-test	Post-test	Gain	G	SEM
1	0.69	0.94	0.25	0.78	0.033
2	0.68	0.91	0.23	0.68	0.046
3	0.73	0.87	0.14	0.52	0.037
4	0.70	0.91	0.21	0.66	0.043
5	0.39	0.72	0.33	0.55	0.035
6	0.56	0.91	0.35	0.78	0.031
7	0.54	0.86	0.32	0.68	0.087
8	0.56	0.77	0.21	0.40	0.057
9	0.52	0.72	0.20	0.40	0.034

Table 1. Mean pre- and post-test scores, gains, and normalized gains, G, for modules 1-9. SEM is the Standard Error of the Mean of G (estimate of error in mean).

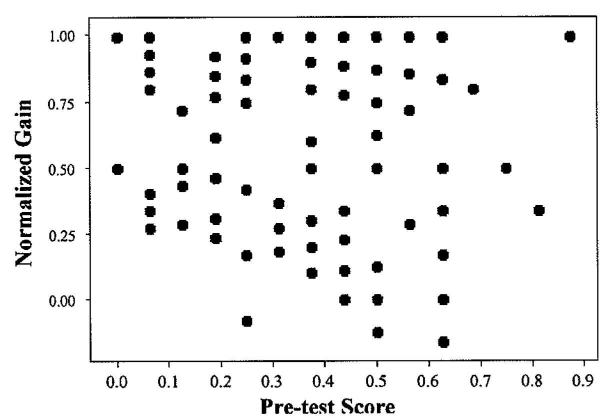

Figure 5. Normalized gains vs. pre-test scores for Module 5 showing significant learning gains regardless of pre-test score.

more in-depth study. From the scatter plot of normalized gains vs. pre-test score for module 5 (Figure 5), it can be seen that nearly all students displayed learning gains and that high gains were achieved by students with high and low pre-test scores. Note that there are many instances of several students having identical scores, so some of the points in Figure 5 represent multiple users.

B. Tracking Patterns of Use: Usage Variations and Markers of Self-Regulation

Since we have no data of students' study habits or thought processes, indicators of self-regulation will be indirect. We argue that one indicator is variation in usage, since different learners will inevitably have different needs for instruction. Thus, the variation in usage across students is of prime interest.

Module	# of LBDs	LBD Mean	LBD S.D.	# of DIGT	DIGT Mean	DIGT S.D.
1 (i)	17	88.0	4.4	9	35.2	12.5
2 (i)	3	85.4	2.1	2	45.3	13.9
3 (i)	12	86.0	9.0	14	39.0	3.5
4 (i)	11	70.7	12.9	9	24.1	4.6
5 (i)	17	61.8	13.8	7	27.9	8.4
5 (c)	17	50.2	11.9	7	22.6	13.0

Table 2. Numbers of LBD and DIGT tutors in modules 1–5, and percentages of students who initiated (rows 1(i)–5(i)) tutors at least once (mean and standard deviation) and completed (row 5 (c)) tutors at least once.

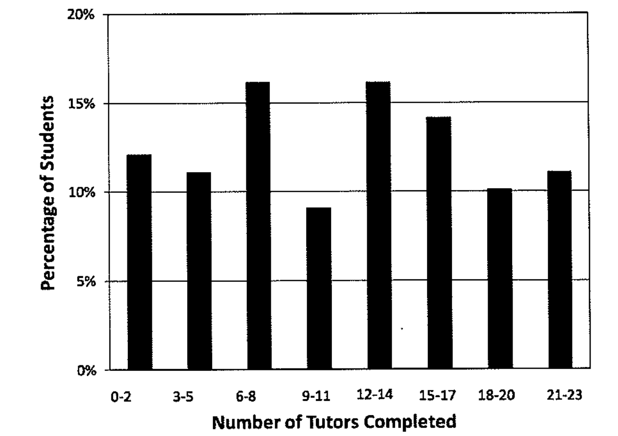

Figure 6. Percentages of students who completed different numbers of tutors in module 5.

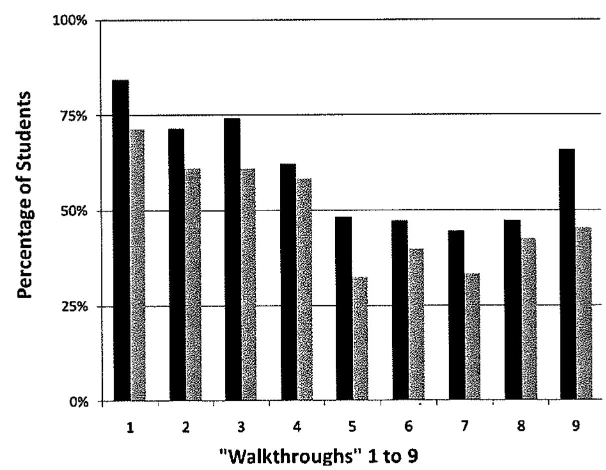

Figure 7. Percentages of students who initiated (black) and completed (gray) each of nine successive "Walkthrough" in modules 1–5.

1) Initiation of LBDs and DIGTs in modules 1–5: The extent to which students avail themselves of tutors appearing in these two forms is of interest. The percentage of students that initiated each LBD and each DIGT at least once was determined. Then, mean and standard deviations over all LBDs and DIGTs in each of the first 5 modules were calculated in Table 2.

It can be seen that usage of LBD tutors is reasonably high at least for the earlier modules, and that it is markedly greater than usage of DIGT tutors. Furthermore, the standard deviations are comparatively low. The difference in usage between LBD and DIGT tutors is statistically significant for each of the 5 modules at the level of $p < 0.001$. Since nearly all students use the LBD tutors, whereas students are more selective in choosing to use DIGT tutors, possibly as it fits their perceived need for additional instruction, DIGT usage may be a marker of self-regulation.

2) Completion of LBD and DIGT in module 5: Tutor completion was monitored for module 5 only. As depicted in Table 2 the completion rate of tutors, represented in row 5 (c), is roughly 80 percent of the initiation rate shown in row 5 (i), for both LBD and DIGT tutors. While there is only modest variation in usage frequency from one tutor to the next (as indicated by the low standard deviations in Table 2), there can be great variability across students. The numbers of students who completed different numbers of tutors out of 24 in module 5 appears in Figure 6. Interestingly, the distribution is not normal, but closer to uniform. Certainly, students cannot be divided into users and non-users; rather usage intensity varies widely. The relation between this variation in usage across students and gains is further discussed in section III C. This more detailed look at module 5 tutors shows that students do not feel compelled to use the materials, and some indeed use them very little. That some students complete nearly all of the tutors suggests that there are likely not enough tutors to satisfy the needs of all students, which is consistent with the survey results reported in Dollár and Steif (2008). In fact, had there been more tutors available, there would likely be even greater variation in usage.

3) Initiation and completion of "Walkthroughs" in modules 1–5: Usage of "Walkthrough" tutors is depicted in Figure 7. There are nine "Walkthrough" tutors distributed across modules 1 to 5. While the rate of usage was higher in the initial modules, the rate stabilized to the point where all "Walkthrough" tutors were completed by at least 30 percent of students. We also determined that the results depicted in Figure 7 do *not* correspond to essentially the same group of students going through all "Walkthrough" tutors. Rather, we found that roughly comparable numbers of students completed between 1 and 9 "Walkthrough" tutors, indicating that nearly all students used them to some extent. Again, this is consistent with an earlier observation that there is a wide variation in the amount of usage across students.

4) Repeated usage of interactive simulation: Study of usage of simulations focused on one important user-controlled interactive simulation from module 5 (EXAMPLE: Equilibrium under Forces Acting in the Same Direction 1) shown in Figure 1. In the simulation, students adjust forces on a bar and then request the resulting motion to be simulated. The goal of the simulation is to recognize that equilibrium requires both translation (force) and rotational (moment) balance. Eighty-seven students worked with this tutor on one to three different occasions. From the log files, we extracted the number of times the "Show Motion" button was clicked (causing a simulation of the motion to which the selected forces give rise) by each student (totaled over all occasions). Figure 8 displays the numbers of students that clicked on the "Show Motion" button different numbers of times. Overall about 75 percent of students clicked "Show Motion" at least once, 40 percent clicked 5 or more

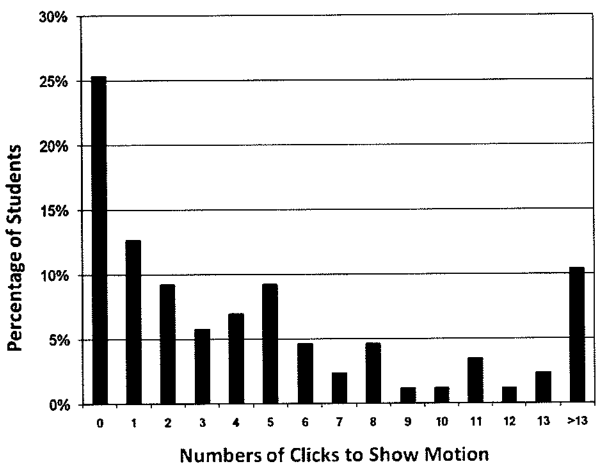

Figure 8. Percentages of students who clicked to invoke motion simulation different numbers of times.

	LBD	DIGT 1	DIGT 2
Initiate	75.7	32.7	27.1
Complete	46.7	31.8	8.4

Table 3. Percentages of students initiating and completing sequential LBD and DIGT tutors.

	DIGT 1	DIGT 2
Users of LBD	40.7	34.6
Non-users of LBD	7.7	3.8
Z	4.37	4.73
p	<0.0001	<0.0001

Table 4. Comparison of percentage of students initiating the subsequent DIGT tutors for two groups: users and non-users of the prior LBD, showing significant difference between groups.

Users of LBD	DIGT 1	DIGT 2
with errors	48.3	37.9
without error	36.5	32.7
Z	1.03	0.47
p	0.305	0.637

Table 5. Comparison of percentage of students initiating the follow-on DIGT tutors for two groups, those who made errors in the prior LBD and those who did not, showing insignificant differences.

times, and 10 percent of students clicked 14 to 33 times. Given the ground rules that usage was not graded, this variation in usage suggests that students use the simulations to varying degrees, possibly in a way that fits their learning trajectory and curiosity.

5) Relation between usage of LDBs and subsequent DIGTs: Since the study of the variation of usage above suggested that DIGT tutors might be used on a self-regulatory basis, a detailed analysis of one LBD tutor and two follow-on DIGT tutors was performed. In particular, we studied if students who interacted with the LBD and hence became familiar with the topic covered in that portion of the courseware were more likely to elect to do the follow-on DIGT tutors. We studied usage of one LBD in module 5 (EXAMPLE: Concurrent Forces 2), which deals with several concurrent forces that keep a body in equilibrium, as well as the two DIGT tutors that immediately follow it. The DIGT tutors dealt, respectively, with choosing appropriate axes for concurrent forces, and with solving a full-fledged problem with concurrent forces. The percentages of the students who initiated and completed these three tutors are given in Table 3.

We then separated students into those who did and did not initiate the LBD, and compared how the two groups used the subsequent DIGT tutors. A statistical comparison of the respective percentages revealed that the difference between the two groups is highly significant (Table 4). This suggests that students work sequentially: they are unlikely to jump to a summative DIGT tutor if they did not already use the previous LBD tutor to practice.

We found that most students who did the DIGT had done the previous LBD. Choosing to do the DIGT may have been conditioned on whether students thought, based on the LBD, that they needed more practice, as well as on whether they still had energy and interest to continue. Therefore, we tried to determine if the performance on the LBD influenced the choice of whether to do the DIGT. To do so we looked at students who did complete the LBD, and split them into two groups: those who made errors in one major phase of the problem, and those who did not make such errors. We speculated that difficulties with the earlier LBD tutor might increase the likelihood of a student re-testing that knowledge on the DIGT tutors. In fact, Table 5 shows that while there is a modest difference in the rates at which these two groups attempted the follow-on DIGT tutors, the difference is not statistically significant. Thus, how students fared on the LBD, at least judged by one measure, had little connection to whether they chose to use the DIGT. It may be that availability of time and energy is the determining factor, or that students chose to use the DIGTs based on an appraisal of their learning progress that is distinct from our LBD performance measure.

C. Relation between Usage and Learning Gains

Whether usage of interactive exercises is correlated with performance is a critical question. In many cases, we found ceiling effects-nearly all students answered many items on the paper and pencil assessment post-tests correctly. In particular, for modules 1 through 4, there were very high means and low standard deviations on the post-test. With the exception of a few individual items, little differences associated with usage would be expected. By contrast, greater variation in performance was found for module 5, so we focused our attention on module 5, which deals with simple cases of equilibrium.

The pencil and paper assessment test corresponding to module 5 contained two problems. The first problem in which forces were concurrent was solved correctly by the vast majority of students. Many aspects of the second problem, which requires consideration of both force and moment equilibrium, were answered correctly and incorrectly by comparable numbers of students. Furthermore, as shown in Figure 6, students differed widely in their usage of interactive tutors in module 5.

1) Correlation between the number of module 5 tutors completed and normalized gain: As a first attempt to relate usage and performance, we computed the standard Pearson correlation coefficient for the number of module 5 tutors completed and the normalized gain. The correlation was relatively low ($r = 0.274$). In a second approach, the class was divided into three groups (almost equal in size), corresponding to whether a low (0–6), medium (7–14), or high (15–23) number of tutors (24 total) in module 5 were completed. Box plots of the normalized gain are displayed for the three groups in Figure 9. In a box plot, the line within the box corresponds to the median, the lower and upper boundaries of the box correspond to the 25th and 75th percentiles, respectively, and the stems reach to the 10th and 90th percentiles, respectively.

Students who completed a medium number of tutors had significantly greater normalized gains in comparison with those who completed a low number of tutors. An ANOVA (Analysis of Variance) indicated that the means of the three groups were significantly different ($F = 4.96, p = 0.009$.) A size effect of 0.64 was computed by dividing the difference of 0.205 between low and medium completers, by the pooled standard deviation of 0.320. The difference in normalized gains between medium and high users was not statistically significant. Pre-test scores of students in the three groups were found to be negligibly different (the pre-test scores were 38 percent, 37 percent, and 36 percent respectively for the three groups). Thus, the lower performance of students who used fewer tutors is not likely to be associated with differing entering ability. Moreover, the number of tutors used did not vary with entering ability, as judged by pre-test score. Why does the increase in tutor usage from medium to high bring less benefits? Two possible interpretations are: (i) there is some duplication in the concepts and skills practiced by the tutors, and (ii) some students who use tutors at least a moderate amount self-regulate, electing to use different numbers of tutors to suit their personal learning needs.

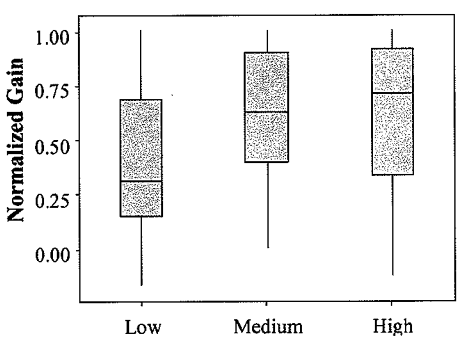

Figure 9. Box plot of normalized gains for groups of students who had completed low (1–6), medium (7–14), and high (15–23) numbers of tutors in module 5.

2) Correlation between tutor usage and student performance on individual questions of the diagnostic test: To determine the presence of a correlation between tutor usage and student performance on individual questions of the diagnostic test, the second problem was considered. The problem featured an L-shaped object to be maintained in equilibrium with fingers supporting it at distinct points, as shown in Figure 10.

This problem had several questions, but two questions in particular, designated here as Q1 and Q2, which pertained to the numbers of additional fingers that were needed to maintain equilibrium and their points of applications, captured the essential lessons of the second part of the module, in which preventing translation in both horizontal and vertical directions and rotation are addressed.

Performance on questions Q1 and Q2 was tracked and compared with tutor usage. We considered only the students who answered incorrectly on the pre-test (68 out of 100 for Q1, and 86 out of 100 for Q2), and we sought to determine if overall tutor usage in module 5 was different for students who answered correctly and incorrectly on the post-test. Table 6 shows that the difference in the mean number of tutors initiated is significant for Q2, but not Q1. The size effect for Q2 is $d = 0.524$, where d is defined as the difference in means relative to the root mean square of the standard deviations.

Finally, analysis was conducted to determine if use of particular tutors was correlated with answering Q1 and Q2 correctly on the post-test. Three tutors were believed to be particularly relevant to the problem on the diagnostic test depicted in Figure 10:

- simulation of a combination of forces that could cause rotation, as well as horizontal and vertical translation, from EXAMPLE: Equilibrium Under General 2D Concentrated Forces 1
- analysis of equilibrium of an L-shaped member balanced with fingers in different positions-an LBD from EXAMPLE: Equilibrium Under General 2D Concentrated Forces 3
- interpretation—an LBD which requests a physical interpretation of the mathematical result of the above analysis of equilibrium of the L-shaped member

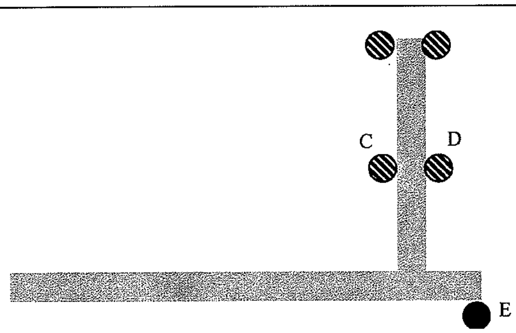

Figure 10. L-shaped object supported by a finger at E and possibly additional fingers at A, B, C, and D, the subject of questions that test concept of equilibrium.

	Q1	Q2
Incorrectly: Mean	9.43	9.14
Incorrectly: S.D	6.71	6.53
Correctly: Mean	11.27	12.57
Correctly: S.D	6.51	6.53
Size effect	0.197	0.524
p (t-test for means)	0.254	0.018

Table 6. Mean and standard deviation of number of tutors completed by students who answered questions Q1 and Q2 on the post-test incorrectly and correctly. Difference in means is significant for Q2.

Students were divided into two (roughly equal sized) groups of users and non-users according to whether they completed the tutors or not (in the case of the simulation tutor, students were divided roughly in half according to whether they ran the simulation fewer than three times vs. three times or more). For each of questions Q1 and Q2, only students who answered incorrectly on the pre-test (the vast majority) were considered. The numbers of students who answered the post-test questions incorrectly and correctly in the non-user (N) and user (U) groups for each tutor is given in Table 7.

For each tutor, a chi-squared test was conducted to determine if the distribution of correct and incorrect answers on the post-test are essentially the same for students who used and did not use the tutor. The resulting values for the probabilities that the distributions are, in fact, the same as given in Table 8. It can be seen that these probabilities are lower in the case of Q2 than Q1. This is consistent with Table 6, where answering question Q2 correctly is shown to be more correlated with overall tutor use than Q1.

Furthermore, the user and non-user groups for tutor types Simulation, Analysis, and Interpretation were successively less likely to have identical frequencies of incorrect and correct answers. Based on a 0.05 level for significance, non-users and users of the Interpretation tutor were significantly different for both questions Q1 and Q2. For the Analysis tutor, differences were significant only for question Q2, and no differences were significant for the Simulation tutor. Of course, the effectiveness of each kind of tutor in producing learning likely depends on the details of its implementation. This merely illustrates that some tutors are clearly capable of producing learning advantages.

3) Gains in module 5 and exam 1 scores in relation to the use of LBD and DIGT tutors: Above, we presented data suggesting that the use of DIGT tutors might signal self-regulation. We note this argument was strongest in the case of the earliest modules (1 to 3) in which there was a high rate of use of LBD's but a markedly lower rate for DIGT. Still, since gains in module 5 exhibited the greatest variability, we compared these gains for frequent and infrequent users of LBD and DIGT tutors in module 5. In particular, we split the class into three groups of approximately equal size, who used LBDs at different frequencies; three groups were also defined with respect to DIGT usage. It can be seen in Figure 11 that LBD use is a predictor of this performance measure, whereas DIGT use is not. An ANOVA reveals that the mean normalized gain of the group with the lowest use of LBD's is significantly different from both groups that used LBD's more ($F = 5.06, p = 0.008$.) There is no difference between different levels of usage of DIGT.

We also considered the scores on class exam 1, which covered material from modules 1 to 5. For each student, we found the number of modules in which at least some DIGT were attempted, and divided the class into six groups, according to whether some DIGT were attempted in 0, 1, 2, up to 5 modules. Box plots are shown in Figure 12 for the six groups; it can be seen that the scores on exam 1 were markedly lower only for students who did not use DIGT in any of the 5 modules. In fact, the mean exam 1 score (71.3 percent) of the 17 students who did not use DIGT was different statistically from the mean exam 1 score (79.7 percent) of the 89 students who did some DIGT ($t = 2.2, p = 0.041$). Again, we would argue that occasionally doing a DIGT tutor signals a

	Simulat.		Analysis		Interpr.	
	N	U	N	U	N	U
Q1 incorrect	20	15	24	11	23	12
Q1 correct	15	18	18	15	13	20
Q2 incorrect	28	21	33	16	32	17
Q2 correct	16	21	17	20	9	28

Table 7. Numbers of incorrect and correct answers to question Q1 and Q2 on post-test, for non-users (N) and users (U) of three tutors: Simulation, Analysis, and Interpretation.

	Simulation	Analysis	Interpret.
Q1	0.335	0.234	0.03
Q2	0.202	0.046	<0.001

Table 8. Probabilities that the user and non-user groups have equal distributions of correct and incorrect answers for three tutors: Simulation, Analysis, Interpretation.

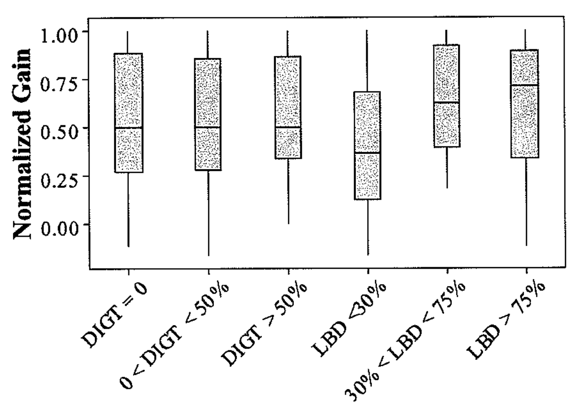

Figure 11. Box plots of normalized gains in module 5 when splitting students in groups according to use of DIGT and LBD tutors.

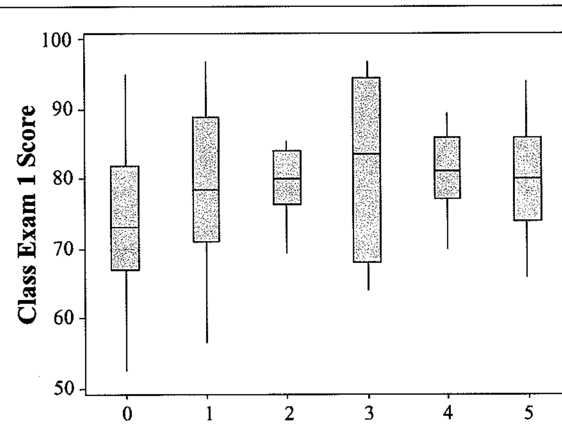

Figure 12. Box plots of Class Exam 1 scores for students who did not use DIGTs at all (0), and who initiated at least one DIGT at the end of 1 to 5 modules. Note that only the group who did not use DIGTs had significantly lower exam 1 scores.

measure of self-regulation. To make an analogous comparison for LBD usage, we compared the 20 percent of students having the lowest rate of LBD usage with the rest of the class and found no difference in mean exam 1 scores. We also tested whether LBD use differentiated among the students who did not use DIGT; there was no difference in exam 1 scores between the sub-groups consisting of more frequent and less frequent LBD users.

Thus, for the class exam 1, which covered material from modules 1 to 5 in most of which there was high LBD use but less DIGT use, students who interacted with DIGT at least a minimal amount performed noticeably better on the exam. By contrast, for module 5 learning gains (based on diagnostic quizzes), LBD use is more correlated with gains. We speculate that the material in module 5 was mostly new for this group of students, particularly in comparison with earlier modules. Not engaging in the initial learning offered by LBD's made it very difficult for students to learn the material in module 5 at all. However, there is again an indication that self-regulated use is important: for learning gains measured by diagnostic quizzes there was little difference between groups that used a medium and high number of LBD's. This is consistent in the sense that little or no usage, which perhaps signals an absence of self-regulation, tended to be correlated with lower performance.

IV. Summary and Conclusions

While different types of courseware can be useful in engineering education, the present paper focuses on courseware featuring many discrete interactive elements. Furthermore, we consider a typical context of higher education in which courseware use is not compelled. Under such circumstances, to take effective advantage of the courseware, users would likely need to, and should be encouraged to, self-regulate their learning. In addition, any assessment of the benefits of the courseware, given the lack of compulsion, should include efforts to gauge just how much the courseware is actually used and whether there are indicators of self-regulation. This prompts the overall research question: what is the relation between learning gains, if present, and overall usage vs. indicators of self-regulated usage?

Students in a traditional statics course were assigned to use nine modules of Web-based learning materials. Learning gains were quantified by paper-and-pencil diagnostic quizzes administered in class, which tested the same concepts and skills as were covered by the Web-based materials. Students received no direct credit for using the courseware modules, although a small credit was given based on students' performance on the diagnostic quizzes. Usage of the intensive, interactive learning activities in the Web-based learning materials was quantified by the analysis of log files in five of the modules. In addition, scores on the first class exam, which covered the material in the first five modules, were monitored for comparison with courseware usage.

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. Although not directly compelled, substantial usage of Web-based learning activities was found. In particular the core learning intensive activities (LBDs) were initiated by, on average, from 62 percent to 88 percent of students, whereas the summative assessment activities (DIGT) were initiated markedly less, by 24 percent to 45 percent. The lower, more selective, use of DIGT tutors, the labeling of these tutors as DIGT (Did I Get This?), which signals their purpose of checking understanding, and the observation that their usage is strongly correlated with completion of the corresponding LBD tutors, all suggest that their use could be a marker of self-regulation.

In-depth study of module 5 (performance on post-test diagnostic quizzes for modules 1 through 4 were uniformly high), revealed a statistically significant lower gain for students who completed fewer tutors as compared to students who completed a medium or high number of tutors. One explanation is that students who completed medium or high number of tutors self-regulated, choosing the number of tutors appropriate to their needs, with comparable outcomes. Gains in module 5 were more correlated with LBD use than with DIGT use. In particular, low scores corresponded to those using few LBD tutors, and ANOVA showed statistically significant difference between low LBD users on the one hand and medium and high LBD users on the other hand. There were not significant differences across low, medium, and high DIGT tutor users. Even though DIGT usage may be viewed, in general, to be a marker of self-regulation, it did not function as such for module 5, perhaps because the material in module 5 was largely new, challenging, and extensive by comparison to the first four modules, (particularly given these students' exposure to elementary aspects of statics in an earlier freshman course). On the other hand, for the first class exam, which covered all of the material in modules 1 to 5 (and more), scores were found to depend on DIGT usage. In particular, those who did not use DIGT tutors at all had markedly lower exam scores; the mean exam scores were comparable for all other students, regardless of the amount of DIGT. This finding is again suggestive of the benefits of self-regulation: being willing to check one's learning—at least occasionally—suggests a user's self-monitoring of learning.

This research has addressed issues that, to the best of our knowledge, have not been raised in the context of engineering courseware, and the following conclusions can be drawn that also point to future research efforts. Methods of *assessing courseware* should go beyond courseware features, learning gains, and student self-reports of effectiveness to include monitoring of actual usage and analyses that relate usage to learning. *Self-regulation* of learning is likely to be critical to successful usage of courseware, and evidence for self-regulation, and means to promote it via courseware design, need to be sought. Clearly, additional work is needed to establish the role of self-regulation in courseware learning, and how students can be encouraged to take better advantage of features in the courseware that facilitate self-regulation. The monitoring of usage should play an additional role in enabling *courseware to be integrated into the larger array of learning activities including those in the classroom.* Data assembled from student online learning activities, if properly interpreted and delivered in a timely manner to affect instructors' actions, could provide powerful insights to both the instructor and the student, potentially allowing students to adapt their subsequent learning, and instructors to adapt their subsequent teaching in the classroom. Courseware design should, increasingly, seek to produce data that meaningfully track student learning, although it is not obvious precisely which data should be assembled and how they should be interpreted for instructors. In the meantime, in the absence of automatic assemblage and interpretation of data on online learning activities, it is important to raise students' interest and willingness to be a partner in the educational process and encourage them to reveal their persistent difficulties to instructors via traditional feedback.

ACKNOWLEDGMENTS

The authors are indebted to the principal developer/programmer on OLI Statics, Ross Strader, as well as to Renee Fisher for her extensive programming contribution. We are grateful to Jamie Lobue for extracting and organizing log data for the analysis. The many suggestions by Marsha Lovett on the data analysis have been very helpful, as has been the thoughtful input of Candace Thille. Support by the William and Flora Hewlett Foundation, by the Department of Mechanical Engineering at Carnegie Mellon University, and by the Department of Mechanical and Manufacturing Engineering at Miami University is gratefully acknowledged.

REFERENCES

Arnold, A., R. Scheines, J. Beck, and B. Jerome. 2005. Time and attention: Students and tasks. AAAI-05: *Educational Data Mining, Technical Report WS-05-02* AAAI Press.

Böhlke, O. 2003. A comparison of student participation levels by group size and language stages during chatroom and face-to-face discussions in German. *CALICO Journal* 21 (1): 67–87.

Brown, A.L. 1980. Metacognitive development and reading. In *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education*, eds. R.J. Spiro, B.C. Bruce, and W.F. Brewer. Hillsdale, NJ: Lawrence Erlbaum.

Campione, J.C., A.L. Brown, and M.L. McConnell. 1988. Metacognition: On the importance of understanding what you are doing. In *Research agenda for mathematics education: The teaching and assessing of mathematics problem solving*, eds. R.I. Chales and E.A. Silver, 93–114. Hillsdale, NJ: Lawrence Erlbaum.

Chapelle, C. 2001. *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge, MA: Cambridge University Press.

Chapelle, C. 2003. *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology.* Amsterdam: John Benjamins Publishing.

Chun, D. 1994. Using computer networking to facilitate the acquisition of interactive competence. *System* 22 (1): 17–31.

Colpaert, J. 2004. Design of online interactive language courseware: Conceptualization, specification and prototyping. Research into the impact of linguistic-didactic functionality on software architecture. *Doctoral dissertation, University of Antwerp.* UMI micro publication number 3141560; also available at http://www.didascalia.be/doc-design.pdf (last accessed July 2009).

Dollár, A., and P.S. Steif. 2008. An interactive, cognitively informed, web-based statics course. *International Journal of Engineering Education* 24 (6): 1229–41.

Dollár, A., and P.S. Steif. 2006. Learning modules for statics. *International Journal of Engineering Education* 22 (2): 381–92.

Educational Data Mining. 1st International Conference, Montreal, Quebec, Canada. http://www.educationaldatamining.org/EDM2008/ (last accessed July 2009).

Fischer, R. 2007. How do we know what students are actually doing? Monitoring students' behavior in CALL. *Computer Assisted Language Learning* 20 (5): 409–42.

Flavell, J.H. 1991. Understanding memory access. In *Cognition and the symbolic processes: Applied and ecological perspectives*, eds. R. Hoffman and D. Palermo. Hillsdale, NJ: Lawrence Erlbaum.

Goodfellow, R. 1999. Evaluating performance, approach and outcome. In *Media, design and applications*, ed. K. Cameron, 109–140. Lisse, The Netherlands: Swets & Zeitlinger.

Hake, R, 1998. Interactive engagement versus traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66 (1): 64–74.

Harris, T.A., and H.R. Jacobs. 1995. On effective methods to teach mechanical design. *Journal of Engineering Education* 84 (4): 343–49.

Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force concept inventory. *The Physics Teacher* 30 (3): 141–58.

Martini, K. 2006. A new kind of software for teaching structural behavior and design. In *Proceedings of the 2006 Building Technology Educators Conference.* College Park, MD.

Nilakant, K., and A. Mitrovic. 2005. Application of data mining in constraint-based intelligent tutoring systems. In *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (896–898).

Open Learning Initiative. http://www.cmu.edu/oli/courses/ (last accessed March 2009).

Opp-Beckman, L., and C. Kieffer. 2004. A collaborative model for online instruction in the teaching of language and culture. In *New perspectives on CALL for second language classrooms*, eds. S. Fotos and C. Browne, 225–251. Mahwah, NJ: Lawrence Erlbaum.

Pahl, C., and C. Donnellan. 2003. Data mining technology for the evaluation of Web-based teaching and learning systems. In *Proceedings of the World Conference on E-Learning in Corporate*, Government, *Healthcare, and Higher Education.* Montreal, Canada.

Romero, C., and S. Ventura. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33 (1): 135–46.

Salaberry, R. 2001. The use of technology for second language learning and teaching: A retrospective. *Modern Language Journal* 85(1): 39–56.

Scardemalia, M., C. Bereiter, and R. Steinbach. 1984. Teachability of reflective processes in written composition. *Cognitive Science* 8 (2): 173–90.

Sherwood, B., and R. Chabay. 1991. Electrical interaction and the atomic structure of matter: Adding qualitative reasoning to calculus-based electricity and magnetism course. In *Proceedings of the NATO Advanced Research Workshop on Learning Electricity or Electronics with Advanced Educational Technology*, ed. M. Caillot, 23–35. Berlin: Springer-Verlag

Steif, P.S. 2004. An articulation of the concepts and skills which underlie engineering statics, In *Proceedings of 34th ASEE/IEEE Frontiers in Education Conference.* Savannah, GA.

Steif, P.S., and M.A. Hansen. 2006. Comparisons between performances in a statics concept inventory and course examinations. *International Journal of Engineering Education* 22 (5): 1070–76.

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

Steif, P.S., and A. Dollár. 2005. Reinventing the teaching of statics. *International Journal of Engineering Education* 21 (4): 723–29.

VanLehn, K., C. Lynch. K. Schulze. J.A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill. 2005. The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, 15 (3): 147–204.

Winne, P.H., and N.E. Perry. 2000. Measuring self-regulated learning. In *Handbook of self-regulation*, eds. M. Boekaerts, P. Pintrich, and M. Zeidner, 531-566. Orlando, FL: Academic Press.

Warnakulasooriya, R., D.J. Palazzo, and D.E. Pritchard . 2007. Time to completion of web-based physics problems with tutoring. *Journal of the Experimental Analysis Of Behavior* 88 (1): 103–113.

Zorrilla, M.E., E. Menasalvas, D. Marin, E. Mora, and J. Segovia. 2005. Web usage mining project for improving Web-based learning sites. In *Proceedings of the 10th International Conference on Computer Aided Systems Theory*. Las Palmas de Gran Canaria, Spain.

AUTHORS' BIOGRAPHIES

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

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

Anna Dollár is an associate 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 teaching has been recognized by many awards including: University Excellence in Teaching Award (IIT), and E. Phillips Knox University Teaching Award (Miami University). Her research focuses on mechanics of solids and engineering education.

Address: MME Department, EGB room 56 E, Miami University, Oxford, OH 45056; telephone: (+1) 513.529.0714; fax: (+1) 513.529.0717; e-mail: dollar@muohio.edu.