

**AC 2009-1753: THE USE AND DEVELOPMENT OF EVIDENCE-BASED SYSTEMS
FOR IMPROVING LEARNING IN ENGINEERING CLASSES**

Jay Martin, University of Wisconsin, Madison

John Mitchell, University of Wisconsin, Madison

Jennifer Welter, Wiley

Sandra Courter, University of Wisconsin, Madison

The Development and Use of Evidence-Based Systems for Improving Learning in Engineering Classes

Introduction

We are interested in using student learning measured via direct assessment as a means of providing evidence for improving learning. This includes enhancing learning both through improvements in instruction and through increased and directed student participation in the learning process.

We are part of the group of developers of concept inventories, in our case, in the thermal sciences. In past work, we have used the results of concept inventory assessment in a variety of ways¹⁻⁴. There were several key findings from our initial studies using the concept inventories: First, there appears to be enormous potential using inventory-type questions to provide measures of student learning. Second, because the results provide a direct measure of student learning, it is a logical progression to next ask how student learning might be improved. Third, the results on the inventory as a whole and on specific topics are a natural means to inform both the instructor and student about where learning is satisfactory, and where additional focus is needed. Finally, we also found that there are many challenges to using a single concept inventory to assess student understanding in a given class or course. The most important challenge arises because of the practical limit on the number of questions in a given inventory. This limits the number of questions for any specific topic, which means that it may be impossible to accurately assess student understanding for all topics in a typical engineering course using a single inventory assessment.

As a result of this past work, we have been engaged in the development of a more comprehensive system for assessment of student learning. Similar to the concept inventories, individual multiple-choice items are the basic components of an on-line assessment system. In contrast to the concept inventories, the system we are developing involves a large number of questions or items (minimum 500 questions per course). Because of the large number of items or questions that have been developed, we can assess student learning at the topic level. The large number of items limits the incentive for students to attempt to memorize specific questions, and in addition, minimizes the possibility that the instructor would “teach to the test.” Finally, because of the number of questions, there is a great deal of flexibility in how the assessment is used. For example, it is possible to provide students with the opportunity to assess their own learning, independent of the instructor.

In the following, we preface discussion of the system with particulars of how we have used concept inventories and what we have learned about assessment from these experiences. Next we will provide details on this system. Finally, we will

describe how we have used this system to date, and what we have learned regarding its effectiveness in measuring student learning.

Prior Work with Concept Inventories

Concept inventories have been used as both formative and summative assessments of student understanding in the physical sciences for more than 30 years. The Force Concept Inventory, which covers six concepts in Newtonian mechanics using 30 questions, is often used as a model for other inventories and is considered to reliably indicate student understanding of the six concepts⁵⁻⁷. Following this model, concept inventories in the three major areas of the thermal sciences (thermodynamics, fluid mechanics, and heat transfer) were developed and used in assessment of student understanding at the University of Wisconsin-Madison. Our experiences in the development and use of these inventories are described in this section.

A major issue with inventory and question or item development is the determination of reliability of the inventory. In this context reliability means “the degree to which an assessment or instrument consistently measures an attribute.”⁸ In particular, we are interested in the development of questions that are a correct indicator of student understanding of the concept, definition, method, or anything that is being assessed by the question. Following Wollack, we estimated item reliability for the concept inventories we developed using item analysis.⁹

To perform item analysis, we gave the inventories to several classes of 50 to 100 students at the end of a semester. Statistical analysis of the results was performed by the Testing and Evaluation Services at the University of Wisconsin. Each class was divided into quintiles based on the overall score, and the percent in a given quintile responding correctly to each question was determined. Figure 1 shows the response curve for a "good" question (No. 8) from the Heat Transfer Inventory that had five possible answers. Question 8 asks:

The thermal efficiency (effectiveness) of a fin is defined as the ratio of the actual heat transfer from the fin to the maximum possible heat transfer. Circle the letter of the statement that best describes the physical meaning of how the maximum possible heat transfer could be achieved.

The possible answers for Question 8 are:

- A The fin has a uniform temperature equal to that of the surrounding air
- B The fin is infinitely long
- C The fin has zero thermal conductivity
- D The convection heat transfer coefficient is infinite
- E The fin has a uniform temperature equal to that of its base

The increase in the percentage of students answering correctly from the 5th to the 1st quintile indicates that the question is probably a good one for discriminating among student responses. The matrix of responses (PROP) shows that all

possible answers were considered by some students and that the majority (55 %) selected the correct answer. The RPBI (point biserial correlation) is an index of the relationship between total scores and whether or not a response was made to that choice. A positive RPBI for an answer indicates a tendency for persons who select that choice to have high scores and for people who do not choose it to have low scores. The RPBI is positive for the correct answer and negative for incorrect choices.

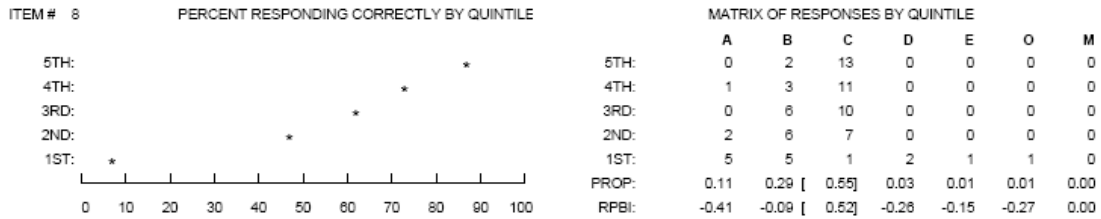
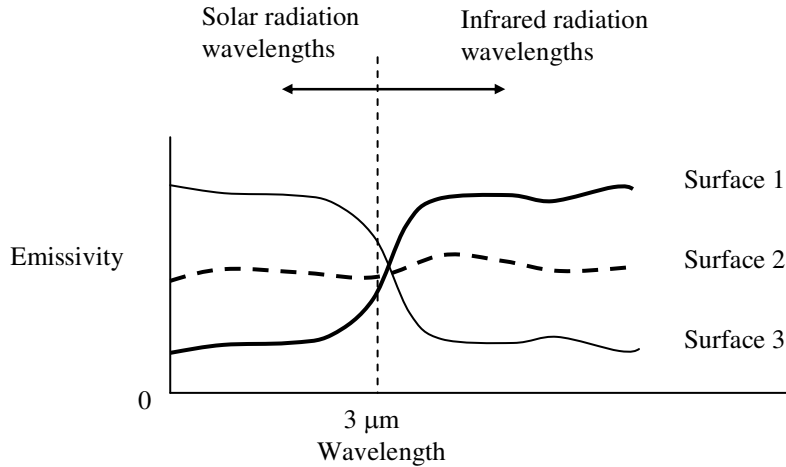


Figure 1 Reliability analysis of a "good" heat transfer concept question.

Although most of the questions had the same general characteristics as Question 8, some were not as satisfactory. Figure 2 shows the item analysis for a "bad" heat transfer question. For this specific question, which experienced heat transfer instructors thought students should be able to answer, the results are basically random and there is no discrimination between the quintiles. There are five possible answers, so the score would be 20 % for random guessing. All of the five answers were selected which shows that even after the course was finished, most students did not get the correct answer. Further analysis established that the question itself was not confusing, but that students just did not understand the idea behind it.

Question 26 asks:

Three surfaces with the same area have spectral emissivities as shown in the figure below. The surfaces are laid horizontally on the ground at noon on a clear day and exposed to sunlight. Most of the solar radiation energy is at wavelengths less than $3\mu\text{m}$ in length, while most of the energy from infrared radiation emitted at temperatures near ambient conditions is at wavelengths greater than $3\mu\text{m}$. What can you determine based on this information?



The possible answers for Question 26 are

- A There is insufficient information to determine the relative temperatures.
- B All three plates reach the same temperature. ($T_1 = T_2 = T_3$)
- C Surface 1 is cooler than 2, which is cooler than 3. ($T_1 < T_2 < T_3$)
- D Surface 1 is warmer than 2, which is warmer than 3. ($T_1 > T_2 > T_3$)
- E Both surface 1 and 3 are hotter than 2. ($T_1 > T_2$ and $T_3 > T_2$)

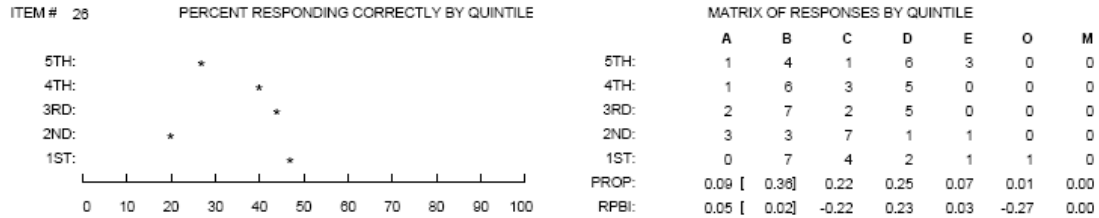


Figure 2 Reliability analysis of a "bad" heat transfer concept question.

The overall summary statistics for the three inventories for several classes for different semesters are shown in Table 1.

The reliability coefficient, also termed alpha, gives an estimate of the internal consistency of the inventories. The higher the reliability, the more confidence one may have that the discriminations between students scoring at different score levels on the test are, in fact, stable differences. A reliability of 0.80 or higher is generally desirable, and the results are close to this limit. Discriminations between student scores can be made with some confidence using these inventories, although efforts should be made to improve the reliability.

Table 1 Statistics for Concept Inventories

Class	No. Students	Mean Score		Standard Deviation	Reliability Coefficient, alpha	Std Error of Meas.
		Score	%			
Thermo S 06	116	21.0	65	4.2	0.69	2.3
F 05	110	19.4	61	3.9	0.66	2.3
Fluids S 06	114	14.9	50	4.1	0.69	2.3
F 05	80	16.9	56	4.7	0.75	2.4
F 04	45	16.5	55	5.8	0.84	2.3
S 04	45	12.4	41	3.5	0.67	2.3
S 04	49	14.2	38	4.4	0.73	2.3
HT S 06	76	15.8	53	5.9	0.84	2.3
F 05	63	18.5	62	5.3	0.81	2.3

The standard error of measurement is an estimate of the error for a given student's score. The standard error was consistently about 2.3 for all of the inventories, and a higher value for the inventory reliability would reduce the standard error. The standard error shows that a student who achieved a score of, for example 15, would be expected to score between 12.7 and 17.3 two-thirds of the time.

The mean test scores were close to 50 % of the maximum. Together with the value of the Standard Deviation, about two-thirds of the students have scores between about 10 and 20 for each inventory.

An open question for the inventories is that of the validity of the questions. This is, are the inventories successful at measuring what we set out to measure, which is conceptual knowledge of the subject? The concept questions were generated by "experts" in the field and reviewed by other experts. Because the material introduced in these courses is by-and-large new to most students, the emphasis was on developing questions that addressed basic understanding rather than misconceptions. The set of questions were generally considered, both by the developers and other instructors, to cover those ideas an undergraduate student would be expected to understand after passing the course. However, it is really difficult to judge that the inventories are "valid" until their reliability is established.

We gave the inventories first within a week or so at the start of the semester (termed Pre) and then at the end of the semester (termed Post). Not all of the instructors participated in the Pre inventory and so the results in Table 2 are not quite consistent with those of Table 1. Summary scores for some of the thermal sciences courses are shown in Table 2.

Table 2 Pre course and Post course scores and gains

Course	Pre Score	Post Score	Gain Score	Post Range Scores	Gain (%)
Thermo S 06	16.4	19.4	3.0	7 - 28	19
F 05	16.3	19.2	2.9	5 - 24	18
Fluids F 05	11.3	17.1	5.7	6 - 28	31
F 05	12.5	16.9	4.5	6 - 28	27
S 06	11.6	14.5	2.8	5 - 22	18
S 06	11.6	15.3	3.6	4 - 21	25
HT F 05	14.4	17.2	2.9	6 - 29	19
F 05	15.4	20.0	4.5	8 - 26	30
S 06	13.4	14.3	0.9	3 - 23	5
S 06	13.9	18.4	4.5	7 - 28	28

In Table 2, the Gain is defined as the actual gain (the difference between the Post and Pre course scores) divided by the maximum possible gain (the difference between 100 % correct and the Pre course score).

We drew a number of conclusions about the performance of students on these inventories. The class average Post score was in the range of 50 to 70 %. The Gains are in the range of 20 – 30 % of the maximum possible gain, which is consistent with the experience reported in the literature. These courses were generally taught in traditional engineering fashion, with formal class lecture periods and an emphasis on problem solving. The gains represent an increase of about 3 to 6 more questions correctly answered at the end of the semester than the beginning.

There is a very large range of correct scores for the individual students in a course. For example, in the Thermodynamics course for spring of 2006, the lowest score was 7 correct out of 32 questions and the highest was 28. This is a range of three or four to one between the highest and lowest scores, and this range between the lowest and highest was consistent for all courses. Since virtually all of the students passed each course, we believe that there are many students with minimal conceptual knowledge of the subject, if these results are valid.

We also used the results to determine those topics in the courses that students had difficulty with. For example, the results for one section of thermodynamics are shown in Table 3. The 32 questions can be put into the five broad categories indicated.

Table 3 Results by topic for Thermodynamics

Topic	Pre	Post	Gain (%)
Properties	51	57	11
Work	72	84	42
Conservation of Mass	70	76	19
Conservation of Energy	49	51	4
Second Law	33	45	18

There are definite differences in the gains for different topics, with a low of 4 % gain and a high of 42 % gain. Both the score and gain for Work are high; apparently students know this topic pretty well. In contrast, the score and gain for the Conservation of Energy are surprisingly low considering how much time is spent on this subject. The inventory can provide information on those topics to concentrate on in future classes, or for which a different approach may be advisable.

Finally, we made an analysis of individual student scores, in which the individual student Post course score was plotted as a function of that student's Pre course score. The results for heat transfer are shown in Figure 3, with similar results for the other classes.

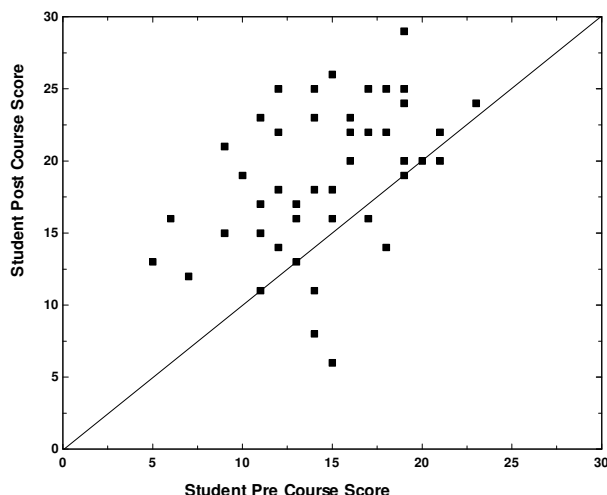


Figure 3 Student Post course score as a function of Pre course score

In general, the Post course scores are higher than the Pre course scores, showing that students in general answered the questions more correctly at the end of the semester than at the beginning. Further, students who had high pre-course scores generally had high post-course scores. However, for all classes, a significant number (about 10%) of the students had lower scores on the Post course inventory than the Pre course inventory, showing that they had negative gains. The "gains" ranged from -6 to $+12$. There are a number of possible reasons for the apparent "negative learning," including guessing on both inventories, "experimental" uncertainty of the inventory, unfamiliarity with the type of question, and actual

confusion about the concepts. Probably, some students did not take the inventory seriously and their results are not valid.

A last analysis was to compare the Post course score to the grade the student received in the course. The grades assigned by the instructors were based on student performance on the traditional problem solving homework and examinations, along with the typical other factors that instructors use in assigning a final grade. The results are shown for the heat transfer course in Figure 4, where an A is equal to 4, a B to 3, and a C to 2.

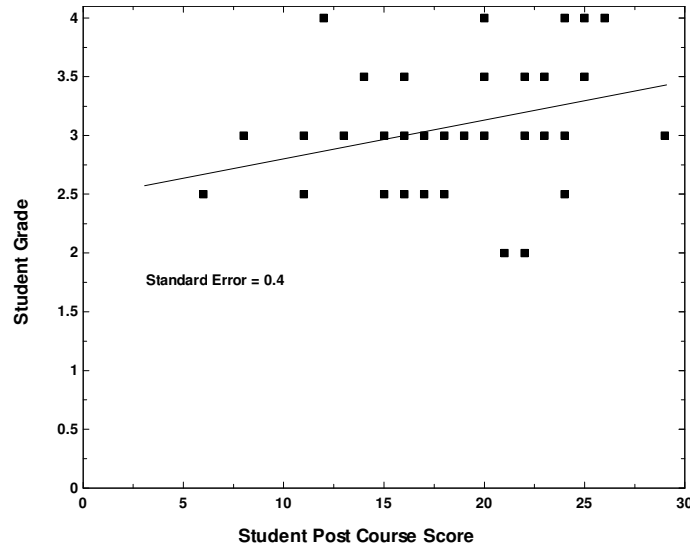


Figure 4 Student grade received as a function of Post score

There is a general correlation between the grade and post-course score. Probably, a student who does well on problems also understands the concepts; the two go together. However, there are some students who get high grades with low scores, and some with high scores and low grades.

Our overall conclusion is that it is difficult to administer and assess the concept inventories in a research university setting. Each class is usually taught by a different instructor each semester, and the semi-autonomous nature of instruction means that different material will be taught in a different manner by each instructor. Although the impetus for using the concept inventories was to meet ABET criteria, instructors were, by-and-large, not supportive of the effort and not involved in administering the inventories. As a result, students were only nominally "rewarded" for taking the inventory and it is possible that many gave the inventory cursory attention. Lastly, some instructors became antagonistic toward the inventory (and the administrators of it) when their student scores were so low. We think it is likely there is as much additional effort needed to develop satisfactory methods to administer the Concept Inventories as there is to develop them.

Description of System

In spite of the difficulties we encountered with the direct use of concept inventories to assess student learning, we saw tremendous potential in the use of assessment methods similar to those used in the concept inventories. In particular, concept inventory assessment indicated that multiple choice questions could provide a reliable means of assessing student learning. Thus we began the development of an assessment question system that would be reliable and valid, that would allow for assessment of student learning, and that would reduce the problems we observed with typical concept inventory assessment.

The assessment system we are developing and currently using is an on-line system, embedded in a course management system. This is an obvious means to enable widespread use of these assessments. The goal of the course management system in use is to manage all (or as many as possible) student and instructor resources, assessment activities, and grading records in one online location, for easy use and reference by both students and faculty. The system includes the following major components:

- An electronic textbook that both students and instructors may access for reading or reference,
- Publisher-provided instructor resources related to the text (Solutions Manual, etc.)
- Assessment questions/exercises for which student answers may be submitted online and automatically graded by the system (including homework exercises from the text, the multiple-choice assessments we have created and reference in this paper, and any additional questions or exercises an individual instructor may choose to create).
- A Gradebook where instructors may track student results on assessment activities both individually and as an aggregate, and by which individual students may track their own results.

As has been mentioned, a large number (500+) of concept questions have been developed for fluid mechanics. The questions developed for fluid mechanics represent the range of concepts and skills a student is expected to understand and apply after an introductory fluid mechanics course, as well as some supporting prerequisite assessment materials for thermodynamics and math skills and concepts. These concept questions are tied to the corresponding sections in the texts.

As individual course and instructor needs may vary, the question bank is planned to be expandable, so that instructors may add their own questions. As new questions are added, it will be desirable to test their validity and reliability. The system is ultimately intended to include a tool for this purpose. The level of

complexity required by the number of questions and their tagging necessitates a delivery system to manage tracking of questions, their related metadata, and student results.

In the current model, to enable ease of use of the system along with the textbook the instructor has chosen, content of individual assessments have been designed based upon a sample syllabus, and aligned to the textbook sections and chapters. So, for example, content related to Fluid Statics (a chapter in most introductory fluid mechanics texts) has pre-designed assessments available. For example, the sample syllabus suggests the section in Fluid Statics on pressures in a fluid at rest be covered in two lectures, with the following pre-designed assessments available:

- 2 Pre-Class Assignments
- Lecture Presentations with 4 in-class questions
- 2 Homework Assignments
- 1 Student Self-Study Assignment consisting of 20 or so questions

The complete bank of questions for an individual content area is larger than what is included in the pre-designed assessments, so that an instructor may have more flexibility in deciding which questions to include in an assignment. For example, the current Fluid Statics content includes 32 questions in pre-designed assignments, but has 46 total questions available for use. The instructor can make the complete bank available to students for their study and self-assessment of their understanding of the material.

An instructor can access the results for individual students and the class (with the ability to drill to the level of detail of which answer a student chose for an individual question), and individual students can access their own results on assessments to better gauge their level of understanding.

By incorporating multiple assessments, there is opportunity for the students to be more engaged with the content, and for the instructor to more easily monitor student performance on both an individual and class-wide basis. Through a combination of greater student engagement, and instructor ability to identify and address gaps in student understanding on an ongoing just-in-time basis, better and more consistent learning can be achieved. Some of the possible ways in which students can be engaged are shown in Table 4.

Table 4 Student Engagement Possibilities

Activity	Student Engagement	Assessment
Pre-Class Assignment	Students read material in order to complete assignment. Student gains basic understanding, or at least has been exposed to the concepts.	Instructor assesses what students already understand, what may be skipped or covered quickly, and what may need more time in lecture.
Lecture/ In-Class Questions	2 nd time seeing the material: Opportunity to improve understanding, or clarify. One question every 15-20 minutes renews student attention span.	In-class questions demonstrate and correct misconceptions. Student group discussion of in-class questions is an opportunity for peer instruction.
Homework	Reinforces and allows students to apply what they learned.	Opportunity for instructors to identify remaining gaps in individual or class understanding.
Student Self-Study	Additional problems with review/study guidance, so students may practice until they feel confident they understand the material.	Students receiving additional practice have recourse, and can still seek instructor guidance in areas of particular difficulty. Directed feedback will be provided to help students study material that they do not understand sufficiently
Exam	Students have opportunity to apply understanding in context of questions covering multiple content areas.	Assessment of student understanding of multiple sub-content areas, and their ability to apply the appropriate concept or skill in a diverse context.

Experience with System and Student Response

As described above, the on-line assessment system is being tested and evaluated in Mechanical Engineering undergraduate classes in Fluid Mechanics. This is the initial test of the on-line assessment system and, the question data set is new and not yet subjected to reliability analyses. Much of our initial experience has centered on using student feedback on use of the system to improve their interaction and use of the system. For example, eliminating mistakes (including

typographical and technical) in the questions have been a major endeavor in which we have enlisted student assistance.

Student response to the assessment system has been evaluated through focus groups held at the midpoint and end of the semester, and through surveys conducted at similar times during the semester. Both the surveys and the focus group activities were also intended to get feedback on the overall on-line course management system, in which the assessment question system is embedded. Here we will provide samples of student response to the assessment question system from the surveys and focus groups. Note that we are providing samples of the student responses that are representative. The methodology we have used is not sufficiently rigorous to allow any quantitative analysis of the student experience at this point. We do intend to do this in the future. Further, the samples we are providing may be confounded because it is likely that students frequently coupled their experience with the overall course management system with their experience with the assessment question system.

For example, the following questions were asked via a survey at the middle of the semester. To understand if the methods used in this course had changed student learning methods, the first question asked:

1. Briefly describe the methods that have helped you learn in this course. Comment on how this course has, if at all, affected your usual learning methods. That is, are your learning methods similar or different than those you use for other courses?

Sample responses included:

- Reviewing concepts
- Concept question and working examples. It makes me change my learning methods. They are a little different.
- It provided me with instant feedback which allows me to break bad habits immediately
- My methods at learning are different for every course. I spend more time in front of my computer for this course
- My learning methods haven't changed much. They are still similar to learning methods for other courses

Some students found reviewing concepts via the assessment question system to be a change in learning methods. Others indicated little effect.

To determine if study habits were influenced, the following questions were asked:

2. Briefly describe your study habits associated with your learning in this course. Comment on how this course has, if at all, affected your usual study habits. That is, are your study habits similar or different than those you use for other courses?

Sample responses included:

- Homework and reviewing lecture notes. I have reviewed concepts more thoroughly. Habits are mainly similar, but have added

- Do attempt to read material before class
- Loved the concept questions before the exam
- I have to have a computer to do the problems or get the questions
- Similar
- I really try to work through as many problems as possible to solidify the concepts
- Read through text to find similar problem; similar to other courses
- My study habits are never consistent. I do homework for this class closer to the deadline than in other classes because it's pretty easy to get done in a short amount of time.
- I mainly use the book to gain understanding for the course.
- I have to study more for this class than other classes and find it very hard to study because it is a very conceptual subject

Again, some students found the assessment questions affected their study habits. Others reported little change.

Regarding whether students have formed a basic understanding of fluid mechanics, the following was asked:

3. At this point in the semester, do you think that you now understand the basic models, vocabulary, order and connections in fluid mechanics?
Please comment on your learning in this course.

Sample responses included:

- I understand the basic concepts, but I'm still feeling way behind on the specifics
- Yes
- Yes, thanks to lecture and homework
- Pretty close to learning/awareness from other classes
- I somewhat have an understanding
- Its made me think more about the equations I use
- Yes, I think overall I have a good understanding (2)
- There are still some basic concepts I am having a hard time understanding

This questions was followed by:

4. At this point in the semester, do you feel that you now have a conceptual framework in fluid mechanics? If so, how would you best describe it?

The responses included:

- Yes (3)
- No (3)
- I understand what we learned in the first third of the semester. However, this middle third has been most challenging
- I had some questions on harder problems that I have addressed to the professor who answered them

- I feel I have a loose conceptual framework. I can see generally what will happen

And finally, we were interested if students would feel that the assessment question system would assist them with remembering concepts:

5. Comment on how, if at all, this course will help you remember the concepts for use in other classes and on-the-job. How do you think your retention of the concepts from this course will compare to your retention of concepts in other courses that you have taken?

Sample responses included:

- Probably not any more or less
- I don't think retention would be very good
- I don't remember the concept for this class as much as other classes
- The fact that we have to state assumptions
- Going through the concept questions before the exam
- I think this will help me remember better than reading alone. The concept questions have helped reinforce material presented in class and in the book.
- Repetition of concept always helps
- Fluid mechanics is pretty intuitive, not a lot of hard formulas/equations to remember like in chemistry, so I'll probably retain lots from this class, but it has nothing to do with how the info is presented
- I doubt that my retention will be much better. The deciding factor in retention for me has been whether or not I sue the concepts at work or for student organizations
- I think my retention will be less because I am having such a struggle to learn this subject

Summary

Experience with the development and use of evidence-based systems for improving instruction in engineering courses leads to the following:

1. There is great potential in using concept-inventory type assessment methods as a convenient and logical means of providing feedback on student learning.
2. Given a sufficient number of items or questions, there are many different ways in which assessment of student learning can be used. For example, the assessment can be used to provide the instructor with evidence on how the class as a whole, or individual students, are doing in understanding course material.
3. A system is being developed that will allow instructors and students to easily carryout these assessments.

4. Testing remains to be done as to the effectiveness of these methods in improving student learning.

References

1. Martin, J., Mitchell, J., Newell, T., "Development of a Concept Inventory for Fluid Mechanics", *FIE 2003 Conference Proceedings*, Nov. 2003.
2. Martin, J., Mitchell, J., Newell, T., "Analysis of the Reliability of the Fluid Mechanics Concept Inventory", *FIE 2004 Conference Proceedings*, Nov. 2004.
3. Foundation Coalition Website: <http://www.foundationcoalition.org/>
4. Martin, J. and Mitchell, J., "Experiences Combining Technology, Assessment, and Feedback to Improve Student Learning in Mechanical Engineering Thermal Sciences Courses, 35th ASEE/IEEE Frontiers in Education Conference, Oct. 2005.
5. Hestenes, David, Wells, Malcolm, and Swackhamer, Greg, "Force Concept Inventory," *The Physics Teacher*, **30** (3), 141-151.
6. Halloun, I.A., and Hestenes, D., "The Initial Knowledge State of College Physics Students," *American Journal of Physics*, 53, No. 11, 1985, pp. 1043-1055.
7. Hake, Richard. 1998. Interactive-engagement vs traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64-74.
8. PALS Guide: <http://www.sri.com/policy/ctl>
9. Testing and Evaluation Services, University of Wisconsin, <http://testing.wisc.edu/WhatDoThoseNumbersMean.htm>, (2009)