Self-Assessment By Students: An Effective, Valid, and Simple Tool?

Sudhir Mehta and Scott Danielson
North Dakota State University

Abstract

This paper gives a progress report of work investigating self-assessment of knowledge by engineering students. Initial results indicate this technique has promise as a tool for assessment of student learning as well as teaching a life-long learning skill. Thus, the technique could be useful for accreditation efforts under ABET 2000. Data from student self-assessment in a senior-level mechanical engineering class were gathered in Spring 1998 (and will continue in Spring 1999). Results from four “in-semester” and a cumulative self-assessment done at the final exam are included.

Introduction

Assessment is typically used for at least one of three purposes: to improve, to inform, and to prove. Tools are available for classroom assessment as described by Angelo and Cross (1993)1 and for program assessment as described by Porus and Johnson (1994)3. Interest in assessment by engineering educators has increased significantly since adoption of ABET’s Engineering Criteria 2000. Meeting ABET 2000 requirements is aided by course learning objectives being clearly articulated and an outcomes-based assessment process used to ensure the learning objectives are being met.

While not as well known as other assessment methods, “self-assessment” may be a tool that could be used at a class level to probe student learning and allow an instructor to tailor classroom instruction to enhance student learning. A comprehensive study on validity of self-evaluation was done by Mabe and West (1982)3. They reviewed 55 studies in which self-evaluation of ability were compared with the measures of performance. The study covered a broad range of activities, clerical, managerial, athletic, etc., but did not address ability of engineering students to self evaluate their knowledge. More recently, Fitzgerald, et al. (1997)2, addressed medical students’ self-assessment of knowledge.

This paper gives a progress report of work investigating self-assessment of knowledge by engineering students. Initial results indicate this technique has promise as a tool for assessment of student learning as well as teaching life-long learning skills.
Methodology

Data collection of self-assessment by students was conducted in two senior-level mechanical engineering classes in the Spring 1998 semester. The classes were “Engineering Measurements” and “Automatic Controls I” with enrollments of 50 and 54, respectively. While data from both the classes were similar, the following specifically addresses only data from the Engineering Measurements course.

Course content was divided into 13 topics or learning areas. During the course of the semester, for four of these topic areas, students were asked to indicate their familiarity (verbal instructions used the word understanding) with the topic before beginning the classroom discussion. Student responses were on scale of one to five with one indicating not familiar at all, three indicating “in-between,” and five indicating very familiar. After all classroom material had been presented, and an “end of topic” quiz taken and returned, the students were asked the same question using the same response scale.

A second set of data were taken immediately after the final examination (which was a cumulative exam). Students were asked to rate their sense of familiarity (understanding) both before and after course coverage with all 13 topics in the class (using the same response scale as before).

Results

The results of the four “in-semester” self-assessments are given in Table 1. The differences between the pre-discussion and post-discussion self-assessment values were all statistically significant (α<0.001). The table’s third column shows the percentage achieved of the maximum possible shift in familiarity. For example, for the instrument calibration topic, a shift of 2.2 (4.1 - 1.9) was achieved out of a maximum possible shift of 3.1 (pre-discussion average of 1.9 subtracted from the maximum possible score of 5) for a percentage gain of 71%. The students’ individual self-assessment ratings of post-discussion familiarity were correlated with their individual quiz performance on that topic. These correlation data are given in the fourth column.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Average rating of familiarity (pre-discussion)</th>
<th>Average rating of familiarity (post-discussion)</th>
<th>Proportion of possible shift achieved</th>
<th>Correlation with quiz performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument Calibration</td>
<td>1.9</td>
<td>4.1</td>
<td>71%</td>
<td>0.60*</td>
</tr>
<tr>
<td>Error Analysis</td>
<td>2.4</td>
<td>3.9</td>
<td>58%</td>
<td>0.36</td>
</tr>
<tr>
<td>Op-Amp (basic)</td>
<td>1.8</td>
<td>3.8</td>
<td>63%</td>
<td>0.34</td>
</tr>
<tr>
<td>Strain Gages</td>
<td>2.7</td>
<td>3.9</td>
<td>52%</td>
<td>0.41</td>
</tr>
</tbody>
</table>
In this case, students had two quizzes over this material. The average score was used for the correlation calculation.

Similar data for the students’ self-evaluation of pre- and post-familiarity for all thirteen topic areas gathered after the final exam are shown in Table 2. The topics are listed in order of coverage during the semester. The four topics on which “in-semester” data were gathered were the first four topics in the course. Again, students’ individual self-assessment ratings of post-discussion familiarity were correlated with their individual quiz performance on that topic. These correlation data are given in the fourth column.

### Table 2
Self-Assessment by Students on Familiarity of Topics (Before and After Class)

<table>
<thead>
<tr>
<th>Topics</th>
<th>Average rating of familiarity before the class</th>
<th>Average rating of familiarity after the class</th>
<th>Proportion of max. possible shift achieved</th>
<th>Correlation with performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument Calibration*</td>
<td>2.7</td>
<td>4.5</td>
<td>78%</td>
<td>0.42</td>
</tr>
<tr>
<td>Error Analysis*</td>
<td>1.8</td>
<td>4.1</td>
<td>72%</td>
<td>0.03</td>
</tr>
<tr>
<td>Op-Amp (basic)*</td>
<td>2.1</td>
<td>4.3</td>
<td>76%</td>
<td>0.13</td>
</tr>
<tr>
<td>Strain Gages*</td>
<td>2.2</td>
<td>4.2</td>
<td>71%</td>
<td>0.06</td>
</tr>
<tr>
<td>Thermocouples</td>
<td>1.5</td>
<td>4.1</td>
<td>74%</td>
<td>0.26</td>
</tr>
<tr>
<td>RTD/Thermistors</td>
<td>1.4</td>
<td>4.0</td>
<td>72%</td>
<td>0.16</td>
</tr>
<tr>
<td>Op-Amp (advanced)</td>
<td>1.7</td>
<td>4.2</td>
<td>76%</td>
<td>0.52</td>
</tr>
<tr>
<td>Linear Motion Sensors</td>
<td>1.8</td>
<td>4.1</td>
<td>72%</td>
<td>0.32</td>
</tr>
<tr>
<td>Opto-electronics</td>
<td>1.6</td>
<td>4.0</td>
<td>71%</td>
<td>0.14</td>
</tr>
<tr>
<td>Angular Velocity</td>
<td>1.7</td>
<td>3.8</td>
<td>64%</td>
<td>0.30</td>
</tr>
<tr>
<td>Encoders</td>
<td>1.8</td>
<td>4.1</td>
<td>72%</td>
<td>0.26</td>
</tr>
<tr>
<td>Data Acquisition</td>
<td>1.8</td>
<td>3.9</td>
<td>66%</td>
<td>0.28</td>
</tr>
<tr>
<td>Flowmeters</td>
<td>1.9</td>
<td>4.0</td>
<td>68%</td>
<td>0.24</td>
</tr>
</tbody>
</table>

*Same topics evaluated during the semester.

### Discussion

The average familiarity of topics before and after class discussion all indicate a positive gain in familiarity with the topic from students’ perspectives. While expected, it is never the less comforting to have validation of the instructor’s effort (something not always self-evident during class). Using Table 2 data, the average percentage of positive shift in familiarity (understanding) achieved is 72%. Both Table 1 and 2 indicate student’s perceived familiarity of the topic before the discussion. Observing this data at the beginning of the topic presentation allows the instructor to adjust the content or pace of instruction on the topic. While not done with these data, a pre-test of the materials could be conducted to allow validation of the student’s self-perceptions.

Table 2 indicates that the familiarity shifts achieved over all topics ranged from 64% to 78%. Longitudinally tracking these data over multiple offering of a course while varying instructional techniques during different offerings of the course would allow the instructor to probe into what worked well, and what did not work well, for each topic with the goal of improved student
learning. This sort of instructional feedback loop is at the heart of the new ABET 2000 accreditation thrust. Obviously, if students do not perceive themselves as understanding a topic very well in the post-discussion data, it behooves the instructor to try and remedy the student’s lack of understanding.

Comparison of Tables 1 and 2 show that students’ perception of familiarity for these four topics (instrument calibration, error analysis, op-amps (basic), and strain gages) had increased by the end of the semester. The differences between the two ratings were significant for the first, third, and fourth topics ($\alpha<0.01$). This increase seems reasonable since the later topics naturally embedded aspects of the earlier material and naturally provided reinforcement of the first four topics. Thus, students should have a better understanding of the material since they would have seen more applications or links between the topics discussed earlier and later.

The highest percentage shift in rating of familiarity was achieved for instrument calibration. As noted earlier, the students were given an opportunity to retake the quiz over instrument calibration. (This was not done on any other topics.) This second quiz provoked another opportunity for students to study and test their knowledge on that topic. This sort of “repeated processing” is key to providing learning gains, however, it requires additional effort from an instructor.

Table 2 includes an example of using this data collection technique to probe effectiveness of instructional technique. The students were assigned to self-study the flowmeter topic using a multi-media module on CD-ROM (developed by Mehta (1996) as part of other work). Table 2 data indicate students’ rated their familiarity with that topic fairly high, similar to levels of the other topics taught in a traditional way. Note that the percentage shift in understanding indicated by the students for the flowmeter topic was actually higher than two other topics actually taught by the instructor (a somewhat depressing observation).

Both Tables 1 and 2 include a correlation of the student’s judgement with quiz performance. The correlation for these two variables for all 17 entries in Table 1 and 2 are positive. However, values are high ($r = 0.41$ to $0.60$) for only four cases and lower ($r = 0.03$ to $0.36$) in the other 13 cases. For the four cases where data were collected during the semester (Table 1), the correlation factors were higher than the corresponding correlation factors based on end of semester ratings (done at final exam, Table 2). Since the familiarity data for these four topics were collected one or two class periods after the topic quizzes were handed back, the quiz scores probably influenced student perception of their familiarity with the topic.

Further, the overall GPA of the students in the instrumentation class were correlated to their entire set of instrumentation class quiz scores. These correlation factors ranged from $0.23$ to $0.55$ with an average of $0.35$. In addition, the correlation factors between student overall GPA and their topic familiarity judgements (post-discussion) for the four “in-semester” data sets ranged from $0.26$ to $0.44$ with an average of $0.35$. However, correlation factors between overall GPA and student rating of topic familiarity taken at the end of the semester ranged from $-0.12$ to $0.25$, with an average of $0.09$. Thus, the end of semester familiarity rating data do not
correlate well with overall GPA while the “in-semester” ratings do. This lack of correlation mirrors the correlation data from the end-of-the-semester familiarity rating with quiz scores.

Correlating student judgement of topic understanding with their quiz scores raises interesting issues. A quiz usually covers only a subset of a topic. Hence, it is possible for a student to have good understanding of the overall topic but still perform poorly on a subset of that topic, or vice versa. Instructional experience also suggests that academic grades or GPA do not have strong correlation with job performance or a successful career.

While it was not done with these engineering students, students can be asked to predict their quiz scores and this prediction correlated with actual scores. Fitzgerald, et al. (1997)\(^2\), did this with medical students and found that, on average, students were highly accurate in predicting their quiz scores. However, they found that students with top 25% scores underestimated their performance by the largest amount and students with the bottom 25% scores overestimated their performance. However, medical students are a select group and this may influence their ability to predict performance.

**Conclusions**

Self-assessment by students can provide valuable insight into the teaching-learning process and its assessment. However, more work needs to be done in terms of the validity of knowledge self-assessment by engineering students. For instance, the students could be asked for self-assessment data before testing on the material and correlation done on those data. The response scale could be changed to allow students a wider spread in the levels of understanding, for example a scale of one to ten instead of one to five. End of class student rating of understanding could be correlated to performance on specific topic areas on the final exam. These and other changes are being planned for implementation during the spring of 1999 and will be reported at the 1999 ASEE annual conference. Also, a longitudinal study of student perception of their understanding compared to job performance in those areas is another possibility.

Asking students for their self-assessment of understanding of content topics or objectives provides useful data. One important use of this data can be to form part of a feedback loop focused on learning-outcomes assessment and improvement of teaching, courses, and curriculum. This use of self-assessment data can be used to demonstrate program evaluation, advising, and monitoring of students (ABET 2000 Criterion 1).

Another interesting issue regarding self-assessment of knowledge is related to professional practice. Practicing engineers often have to make judgements regarding their knowledge or ability to perform various engineering tasks. These judgements have importance in both legal and ethical issues during professional practice. Thus, there is value in teaching engineering students about their ability to correctly self-judge knowledge. By regularly having students judge themselves and providing feedback on the accuracy of that judgement, engineering educators have an opportunity to demonstrate that they are meeting the ABET 2000 Criterion #1, “a recognition of the need for, and an ability to, engage in life-long learning.”
References


Biographies

SUDHIR MEHTA
Sudhir Mehta is a professor of Mechanical Engineering at North Dakota State University. He was named the 1997 North Dakota Professor of the Year by the Carnegie Foundation. His areas of interest are engineering education research, instrumentation, controls, robotics, design optimization, and machine vision. He has developed 2 CD-ROM’s containing hypermedia based instrumentation and communication resource modules, He has also developed innovative techniques for active learning, collaborative learning, and quick assessment. Dr. Mehta received the Carnot Award for the best teacher of the year, four times, from the students of Pi Tau Sigma Society. His e-mail address is mehta@badlands.nodak.edu.

SCOTT DANIELSON
Scott Danielson is an assistant professor at North Dakota State University. He was the chair of the Engineering Technology Department and directed its Aero Manufacturing Engineering Technology program. He is currently a faculty member in the Industrial and Manufacturing Engineering department. Other responsibilities at NDSU have included teaching courses in the Mechanical Engineering and Applied Mechanics Department and managing the Robert Perkins Engineering Computer Center. Dr. Danielson received the College of Engineering and Architecture’s Teacher of the Year Award for the 1995-1996 year. His research interests include effective teaching and engineering applications of geographic information systems. Before coming to the University, he was a design engineer, maintenance supervisor, and plant engineer. He is a registered professional engineer. His e-mail address is sdaniels@plains.nodak.edu.