AC 2009-1755: ERROR TRACKING: AN ASSESSMENT TOOL FOR SMALL-ENROLLMENT COURSES

Anne-Marie Lerner, University of Wisconsin, Platteville

Anne-Marie Lerner is a first-year assistant professor at the University of Wisconsin - Platteville collaborative program located at the University of Wisconsin - Rock County. Her research interests include assessment, engineering education, K-12 outreach, speech processing, and semiactive vibration control. She received her PhD in mechanical engineering from Georgia Institute of Technology in 2008.

© American Society for Engineering Education, 2009

Error Tracking: An Assessment Tool for Small-Enrollment Courses

Introduction:

A program undergoing ABET accreditation must institute a procedure of assessing student outcome[1]. This is often done in class-by-class analysis. Instructors teaching large-enrollment classes have enough students that statistically significant assessment data can be collected without significant hardship to the professor. For small-enrollment classes, the same assessment data may not provide any meaningful information, as there may or may not be enough data points to derive statistically significant conclusions. A novel method of assessment comprised of categorizing and tracking individual errors is presented and discussed in this paper as a means of assessing students in small-enrollment classes.

Background:

A method of error tracking was developed in order to assess students in small-enrollment classes at a university collaborative program at another location. The collaborative program began in 2005 in a city; two electrical engineering faculty members and one mechanical engineering faculty member work on-site at the host university and offer primarily night and evening classes to place-bound students. The electrical engineering program was started up first, where courses began being offered in 2006; mechanical engineering began in the fall semester of 2008. While students stay on the host campus to take their courses, they receive a degree-granting institution's degree upon completion of their coursework. The host university courses must be offered in such a way so as to provide an equivalent education to the courses offered directly on the degree granting institution's main campus. Furthermore, to maintain ABET accreditation, the assessment must show an equivalent education.

This program is designed to provide educations to place-bound, nontraditional students, many of whom cannot support a full-time student load. Furthermore, many students who enter into the program may have to take classes such as college algebra or even more remedial math. While the overall engineering program has over sixty students enrolled in it, with over twenty listed as mechanical engineering students, very few students have the necessary prerequisite math to be able to directly enroll for mechanical engineering courses. As a consequence, for the first semester of mechanical engineering offerings, thermodynamics had an enrollment of six students, and engineering materials saw an enrollment of three students. The average enrollment of an on-campus offering of thermodynamics is 25 per class, whereas engineering materials has an average enrollment of 43 students per class offering.

Assessment Methodology: Error Classification

Students' work was evaluated for root errors during the grading process. A root error is defined as the primary incorrect thought that leads to an incorrect answer, although more than one root error may be identified for an individual problem. A record of that root error was made by assigning the error to one of the following ten error classifications:

Error Significance	Error Code
Minor Issues	Math Error Missed Detail
Intermediate Issues	Incorrect/Lack of Unit Conversion Sign Error (of significance) Partial Completion
Major Issues	Lack of Thought/Comprehension of Problem Correct Analysis, Incorrect Conclusions Incorrect Formula Applications Inappropriate Assumption Got Stuck – Failed to Use Formula

Table 1: Ten error codes and their significance

The ten error codes are placed into three groups in order of the severity of the error. For example, a math error is considered a minor error because it is not indicative of a lack of comprehension of the concept at hand. On the other hand, if a student got stuck and failed to find the correct formula to use, this is a sign that the student does not sufficiently understand the issue at hand. While errors in the "Major Issues" category are significant and deserve attention, those in the intermediate and minor issue classifications can be indicative of a serious problem if they occur habitually.

Grouping errors into classes of different severities is not a novel idea; Gagne's learning theory divides learning into eight different classifications, with the assertion that higher-order learning classifications cannot be achieved without lower-order learning first having occurred[3]. However, this method was designed for engineering-specific homework evaluation, and in particular examines the mistakes students make during the learning process.

This error assessment is meant to be conducted on an on-going basis, where the instructor uses the feedback in real-time to adjust lesson plans to address any deficiencies in students' understanding. Appropriate actions for error assessment depend on which group the particular error code belongs to. A minor issue error is expected to pop up periodically throughout a student's work, and should not be concerning unless there is a trend of fundamental lack of mathematical understanding as described by Buechler[2]. In such a case, error tracking should be used where the subject of the error is specifically the math deficiencies of concern, such as derivatives, matrix algebra, or Laplace transforms. Intermediate errors are the ones that transcend the specific subject of study and that all instructors are working to minimize; the incidence of these errors should decrease as the semester progresses. An increase in these errors would be an indicator that this is a subject that should be reviewed. A major issue incident, on the other hand, shows that the student has a significant gap in knowledge.

Different tracking methods can be used to evaluate student learning. Intermediate errors can be tracked over time to assess whether students are steadily improving in their work in these arenas. In this case, specific knowledge is not tracked, only the incidence of these intermediate errors. A second method can be employed by tracking specific gaps in knowledge over time, where in this case the specific skill or knowledge set is tracked over time, while the precise error class is not. The next section gives examples for each of these types of tracking.

Assessment Implementation: Two Examples of Error Tracking

Error is tracked against assessment incident, so that performance over time can be measured. For this assessment method, an error ratio is calculated such that for every assessment opportunity, the number of errors of interest (which can be an error class, an error type, or a specific error) divided by the number of opportunities to make an error within the same assessment opportunity. The error ratio is calculated as

$$r_e = \frac{e_i}{e_T},$$

where r_e is the error ratio, e_i is the number of errors of interest measured at the n^{th} assessment time, and e_T is the total number of opportunities for the errors to be made, measured at the n^{th} assessment time. The total number of errors, or e_T, is taken to be the number of problems worked within an assessment period. Therefore, if six students worked five problems apiece, the total errors possible would be 30. Generally, $e_i < e_T$, since there is generally one overriding error that causes problems. However, for an individual problem it is possible to have more than one primary error, and therefore it is possible that $e_i \geq e_T$. Error here is tracked as a fraction so that error incidences can be compared between incongruous assignment lengths, and also so that students with very low absolute rates of error commission (ie: strong students) didn't skew the percentages with a rare error.

By examining the change in the error ratio discussed above, the overall errors within a class are tracked and compared to a qualitative assessment of the course, as well as two assessment methods. Figure 1 shows the total error ratio for each assessment opportunity. The fifth, tenth, and fourteenth assessment opportunities represent midterms and a final exam, while the remaining eleven assessment opportunities represent graded homework assignments.

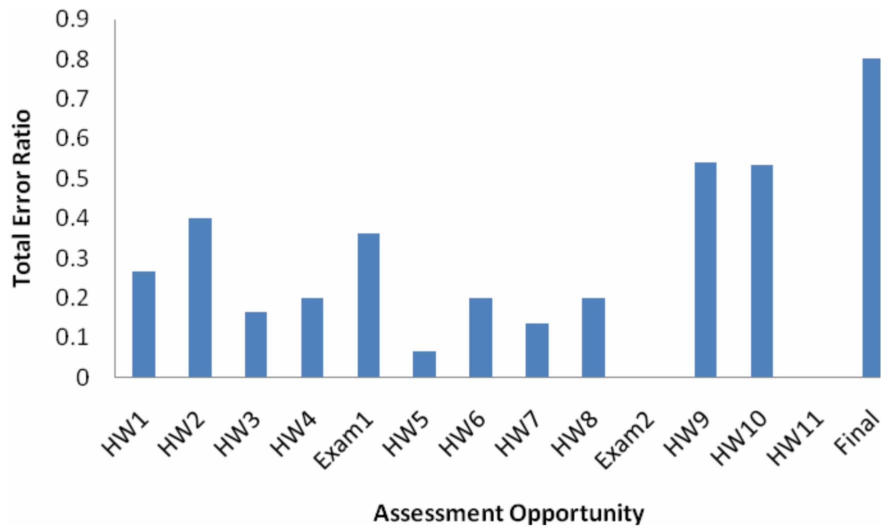

Figure 1: The total error ratio is plotted against chronological assessment opportunities for an undergraduate thermodynamics course.

There were two assessment opportunities where every student answered each question correctly, exam 2 and homework #11. These two events were particular anomalies. Exam 2 was a little too similar to the practice exam, and since students had prepared well with the final, they were able to master the exam. Homework #11 was supposed to be assigned (and the material that pertained to this homework was also to be taught) the day that classes were cancelled due to inclement weather; as a result, the due date was delayed until there was not enough time for them to study from it for the final. As a compromise, and to assist with their studying, they were given select answers. The result was that the students came to office hours in overwhelming numbers to receive assistance, and worked until their solutions matched the given select answers.

Figure 1 can be used in conjunction with qualitative observations about the class to determine where in the semester course content is more challenging for students. For example, the overall trends of Figure 1 show that error ratios averaged around 30% for the first third of the class, decreased to about 20% on average for the middle third, and then increased to around 50% for the remaining third of the class. This corresponds to the qualitative observations that students struggled to understand and internalize mass balances and the 1st law of thermodynamics in the first third of the class. The second third of the class was spent applying the 1st law to various devices (pumps, condensers, nozzles, etc.) and since this involved slight application variants upon the first law, students tended to master this material easier. They also conceptually understood the second law of thermodynamics. Homework #9 corresponds with the introduction of entropy, and students struggled with this concept. The total error ratio reflects this difficulty in concept mastery. For an instructor, the total fraction error is a useful tool to confirm qualitative observations about the course after it has been completed, and can then be used to plan how future offerings would be taught.

While the error ratio can be used to examine overall errors, it can also be used as assessment methodology tools. The first method involves tracking specific intermediate errors that should decrease over the semester, regardless of content. This method will be referred to as "Common Errors Tracking". The second method involves tracking specific knowledge gaps over time, regardless of error type, and is referred to as "Specific Knowledge Tracking".

Figure 2 shows an example of common errors tracking for a thermodynamics class with six students, where the error ratio is plotted against assessment opportunity. The two common errors tracked were unit conversion and sign errors. The error ratio committed by the class is charted against the assessment opportunity. While most of these assessment opportunities arose in the form of homework, the fifth and tenth assessment opportunities both represent midterm exams, where students were allowed one page of notes, and the 14th assessment opportunity represents the final exam.

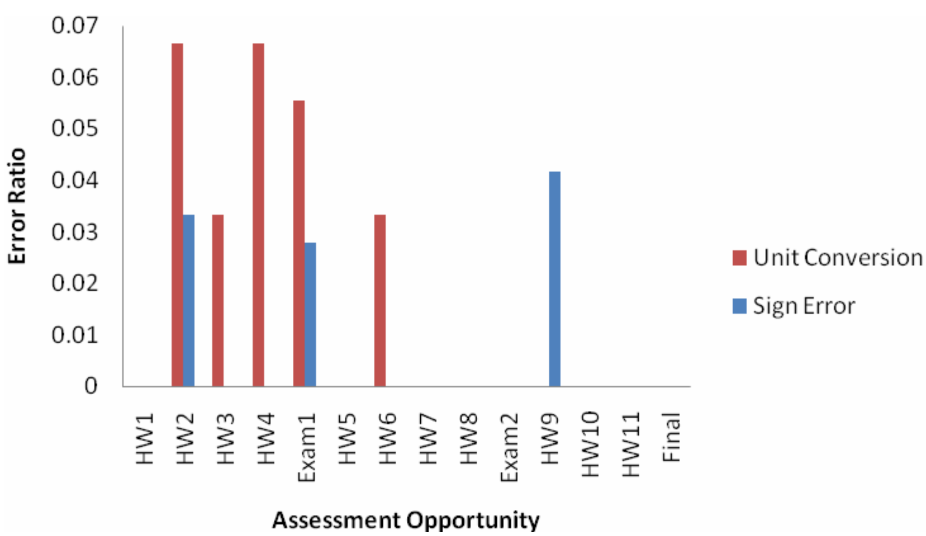

Figure 2: Error ratios are shown against assessment opportunities for unit conversions and sign errors.

Figure 2 shows that incidences of unit conversion error most frequently occurred at the beginning of the semester, tapering off completely by the eighth assessment period. The incidences of sign errors also decreased over time, although sign error incidences occurred through the eleventh assessment. From Figure 2 it can be seen that students improved their performance in both of these areas over the duration of the semester. It remains to be seen if this performance improvement can be sustained over semester breaks and into other classes.

While Figure 2 shows the incidence of common errors versus assessment opportunities, Figure 3 shows the incidence of a specific knowledge error. The specific knowledge error in this example is that of plotting pressure versus specific volume. While students were able to find the endpoint values for each stage, the tendency was to draw straight-line connections between the various state endpoints rather than the appropriate curves. The context for this particular error is discussed, along with the educational interventions attempted, in describing Figure 3.

In homework #1, students were asked to draw a *p-V* diagram of a three-stage cycle. Two stages were either constant-pressure or constant-volume conditions. The third condition was an ideal gas expansion, or $pV = constant$. All students were successful in identifying the three paired pressure-volume coordinates, but then erroneously connected the dots with straight lines. In the next class period, this particular issue was discussed, example problems were given, students worked in-class problems to address this issue, and then homework #2 included a similar problem for reassessment. Barring a few math errors, all students were successfully able to master this concept. Most students remembered how to do this correctly for the first exam, but in the absence of any reinforcement throughout the semester, students fell back to connecting the dots with straight lines rather than considering specific *p-V* relationships.

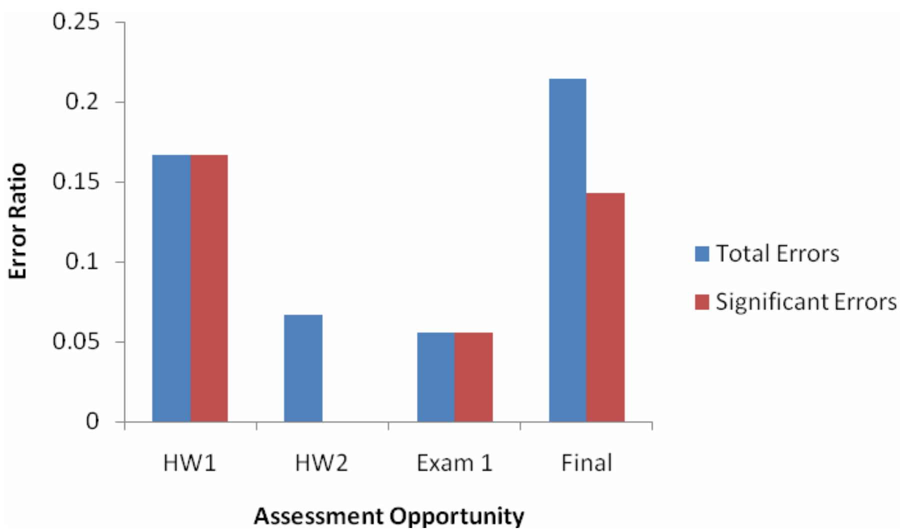

Figure 3: Error ratio against assessment opportunities for a specific knowledge error.

Figure 3 shows two different types of error ratios; total errors are plotted, as well as significant errors. Referring back to Table 1, a significant error in this context was designated as a partial completion, lack of thought/comprehension, or inappropriate application of a formula. This highlights the somewhat subjective criteria for judging an error. In this instructor's estimation, connecting the dots with straight lines is a sign of a student not thinking carefully about a problem, hence a lack of thought/comprehension error designation instead of an inappropriate assumption designation. The instructor in this case is attempting to judge motivations behind the errors. While it is expected that each individual instructor may attribute the same hypothetical error to different root error classes, there will at least be a consistent frame of reference for an individual instructor.

While error assessment was implemented successfully in a lecture-based thermodynamics, particularly in analyzing homework and exams, and with limited success for a lab-based engineering materials course, it is anticipated that this methodology could be expanded to encompass assessment for laboratory reports, presentations, and group work. In the next section, future implementation plans are detailed along with anticipated challenges.

Error Assessment in the Future: Proposed Implementation

Error assessment was implemented successfully in a sophomore-level thermodynamics class. It was also implemented to a less-successful degree in a sophomore-level engineering materials class. The difference in success lies in how the error assessment methodology should be implemented in different data collection methods. For example, error assessment for homework and exam problems asking for quantitative calculations are very straight-forward, hence the success in the thermodynamics class. A different strategy must be implemented when applying error assessment to qualitative homework responses, lab reports, presentations, and group work. Furthermore, this method should be studied in large class sections to allow for a statistical validation of this methodology. This section addresses proposals for implementing error assessment in these cases, and also how to normalize error assessment across instructors.

Error Assessment for Labs and Qualitative Responses

When evaluating a qualitative response from a student, such as an essay question or a lab report, some of the error assessment categories are no longer relevant. For example, a math error would be unexpected in response to the question, "What is the difference between normalizing, full anneal, and spheroidizing?" On the other hand, a missed detail is relevant, as is partial completion, and lack of thought/comprehension. In that way, this short-answer question can be subjected to an error analysis. However, since so few of the error codes are applicable in this case, it makes sense to tag the problem as having minor, intermediate, or major issues.

In the case of a lab report, error analysis as discussed in this paper should restricted to the domain of technical content, in which case the entire error code catalog may be used to track mistakes. A second set of error codes must be developed to encompass grammatical, stylistic, and formatting issues that can be used to evaluate lab reports. The original error codes were developed throughout the semester as a way to class errors that were seen repetitively. The same method would be employed for evaluating the nontechnical components of a lab report. This method can then be evaluated for usefulness in the assessment process.

Error Assessment Normalization Against Different Instructor Analysis

It is expected that different instructors would evaluate errors differently. One instructor's "lack of thought/comprehension" might be another instructor's "inappropriate assumption". However, while the specific error codes may be disparate across instructors, the significance would be expected to remain constant. Intermediate errors, which were tracked by error code in Figure 2, can be grouped together and compared across instructors for the duration of the class, or evaluated in lump-sum at the end of the semester. It would also be useful to see if there is any correlation between error assessment and any student evaluations for the instructors.

Validation Through Implementation in Larger Class Sizes

The method of error assessment, as discussed in this paper, has to date only been implemented in classes of such a small size as to not yield statistically significant results. Implementation across a larger class size would provide an opportunity to evaluate this assessment method and draw statistically significant conclusions about this methodology.

Conclusions

A method of error assessment was developed for use in small-enrollment classes at the degree granting institution's collaborative program at host university location. The error assessment method involves placing every error into one of nine categories. These nine categories were further lumped into three groups, depending on the severity of the error type. Errors in general that were tracked over the course of the semester agreed with the instructor's qualitative assessment of the students' understanding. This analysis is not useful for real-time intervention, but is useful as a post-semester analysis for future course changes. Errors could be further assessed by error group over time. Students' understanding of a particular knowledge set was also studied by looking at overall fractional error incidence associated with assessment of that particular knowledge base over time. This particular implementation of error analysis was successfully used for real-time intervention with knowledge gaps. Future expansions of the error analysis technique will involve adapting the technique for analysis of lab reports and qualitative question analysis, as well as ways in which to make instructor-by-instructor analyses meaningful.

References:

[1]Engineering Accreditation Commission, "2008-2009 Criteria for Accrediting Engineering Programs". Nov. 3, 2007.

[2]Dale N. Buechler. "Investigating the Mathematical Background of Engineering Graduates to Improve Student Retention". Proceedings of the 2004 North Midwest ASEE Section Conference, 2004

[3]Tom E. Lawson. "Gagne's learning theory applied to technical instruction". Training and Development Journal, April 1974 Vol. 28(4) 32 - 40.