FINAL REPORT COVER SHEET
R. PATRICIA WALSH GRANTS IN THE SCHOLARSHIP OF TEACHING AND LEARNING, 2012-2013

Name(s): Matt Sniawski & John Dionisio

Department(s): Mechanical Engineering; Electrical Engineering & Computer Science
Design and Pedagogical Value of a Standards-based Grading System for Undergraduate STEM Education

Abstract: Most science, technology, engineering and mathematics (STEM) higher education instructors utilize summative score-based grading systems to assess student performance. However, these traditional systems typically do not directly assess student development towards achieving course objectives. Standards-based grading is an alternative approach to assessment of student performance and learning. It involves the direct measurement of student development towards achieving specific course objectives. Student development is tracked throughout the duration of a course using a standards development report rather than assigning one-time individual scores to student assignments. Final course grades are then determined based on their overall development towards achieving course outcomes. One of the major benefits of standards-based grading is that it provides clear, meaningful and personalized feedback for both students and educators regarding student learning. Although it has gained popularity at the K-12 level, there have been no studies to date that analyze the effects of standards-based grading on undergraduate STEM education. Pilot studies were conducted in multiple STEM courses to assess the pedagogical value of standards-based grading and its impact on student cognitive and affective behaviors. Affective behavior was measured by assessing changes in students' self-efficacy and the value they place on standards-based grading. Cognitive behavior was measured by assessing students' epistemological beliefs of STEM knowledge. These studies of standards-based grading suggest that standards-based grading can have a beneficial effect on STEM higher education, but may have varying value based on the particular type of course.

Keywords: assessment, grading, STEM education, learning outcomes

Awardee(s): Matt Sniawski John Dionisio 5/21/2013

Chair(s): Nader Sarei Ray Teal 5/21/2013

CET Chair: Signature 1-13-12 Date

Please attach a description of the results, their dissemination, and all relevant supplementary material as described in the guidelines.
Results

Work Accomplished
This project implemented standards-based grading in selected engineering and computer science courses over the 2012–2013 academic year with the goal of assessing its pedagogical value in undergraduate science, technology, engineering, and mathematics (STEM) education and learning how it impacts student cognitive and affective behaviors. The project tested the hypothesis that the standards-based grading system will produce statistically significant changes in both student affective and cognitive behaviors regarding STEM concepts and learning. Work involved the design and implementation of the grading system, the creation of instruments to measure student response to the system, and an analysis of the results after the courses ended.

Methods Used
We deployed the standards-based grading system in a total of 8 undergraduate STEM courses during the 2012–2013 academic year: CIVL 200, CMSI 370, CMSI 486, CMSI 371, CMSI 387, ENGR 200, MECH 483, and MECH 484. The work builds on pilots done in academic year 2011–2012. A qualitative evaluation of the specific implementation of standards-based grading was done between semesters, leading to incremental revisions to how outcomes were stated and scored. The overall approach of outcome-level granularity without conflation into a single cumulative score (except for the final grade) was retained throughout, however.

Affective behavior was measured by assessing changes in students’ self-efficacy and the value they placed on standards-based grading. Self-efficacy, or an individual’s confidence about their ability, is shaped by experiences, persuasions, and physiological states (Bandura, 1986). Cognitive behavior was measured by assessing students’ epistemological beliefs of STEM knowledge, or the beliefs about the nature of knowing and learning. This added variable enhanced our analysis of cognitive gains through the identification of how naïve or sophisticated the students’ understanding is (Hofer & Pintrich, 1997; Schommer, 1990).

Changes in both self-efficacy and epistemological beliefs were evaluated using pre- and post-course surveys. The instrument that was developed to measure self-efficacy was a modification of a previously validated instrument (Carberry, Lee, & Ohland, 2010). The measure of epistemological beliefs was recorded using a modified version of the Epistemological Beliefs Assessment for Engineering (EBAE) (Carberry, Swan, & Ohland, 2010). Value placed on standards-based grading was measured using post-course surveys that we also developed.

Main Findings
Overall, we found that standards-based grading represents an alternative to traditional score-based grading. It reflects student performance in a simultaneously more transparent and
granular manner than cumulative score-based grading. For engineering courses, the system was found to integrate well with ABET (Accreditation Board for Engineering and Technology) assessment.

Not surprisingly, we found that the selection of outcomes/standards for the course was instrumental in our qualitative view of how well the grading system reflected student learning. With each piloted course, we found that:

- The type of outcome/standard may call for different final measurement approaches. For example, an outcome that represents a “best practice” that must be accomplished for every assignment may be best finalized as an average. Outcomes that are achieved by an increasingly advanced sequence of activities may be best finalized by the most recent recorded proficiency given to that student.
- Some outcomes may need to be weighted to reflect instructor priorities.
- Standards-based grading is orthogonal to the assessment of group or collaborative work. That is, group assessment tools such as CATME (Comprehensive Assessment of Team-Maker Effectiveness) remain valuable in courses that have team assignments or projects.

As future courses are adapted for standards-based grading, or previously adapted courses are offered again, we anticipate continuing refinement of how the outcomes/standards that anchor this grading system are best stated and evaluated.

With regard to implementation, we found that providing feedback along the lines of the defined outcomes/standards helped determine points of strength and improvement for the class. This resulted in empirical guidance for where additional time or focus must be spent. The logistics of distributing feedback could be improved, however—the structure of the grading system was sufficiently different that existing tools such as Blackboard required significant adjustment or effort. Most courses were tracked using specially designed spreadsheets. For computer science courses, a version control system that students already used for their code was also used for disseminating feedback and standards development reports. This allowed them to submit course work and see the feedback on that work using the same file service. Overall, however, a case can still be made for a software application that is expressly designed to support standards-based grading. We have begun to pursue this objective through an internal Academic Technology Grant.

Based on our work, standards-based grading appears to be a viable approach for STEM undergraduate courses in general, although may have varying value based on the particular type of course. In particular, courses with a strong group or collaborative element, as well as courses that are based on term-length projects (e.g., engineering design; software development) need additional work and study.
**Assessment Results**

Our implementations of standards-based grading positively impacted students’ affective and cognitive behaviors. Their self-efficacy increased and epistemological beliefs became more sophisticated. Students saw high interest, attainment, and utility value in the grading system while not costing them additional valuable time and effort in order to learn the material.

Figure 1 illustrates changes in self-efficacy for one of the pilot courses in this project, selected for its sample size and small number of measured outcomes. A correlation between improved self-efficacy and the final grade received by a student was seen in some engineering courses, but not in computer science. Further exploration of this result may be worthwhile but will need larger student cohorts than were available for this project, particularly in computer science.

For courses that had a sufficient number of students for control and experimental groups, self-efficacy gains pre- and post-course in the experimental group were statistically significant, but not in the smaller control group. There were no observed differences in value of the grading system between the experimental and control groups.

![Figure 1. Changes in mean self-efficacy for MECH 212. Students were asked to rate their confidence in performing the tasks on the left, with 0 = cannot do at all, 50 = moderately certain can do, and 100 = highly certain can do.](image-url)
Figure 2 shows post-course student value of the grading system. Students overwhelmingly agreed with value statements expressing interest, attainment, or utility with regard to standards-based grading. Conversely, they largely disagreed with value statements expressing potential costs or disadvantages associated with standards-based grading (marked with an asterisk).

Figure 2. Post-course student value of the grading system (aggregate; 63 total responses distributed over 5 courses for which the survey was administered).
Project’s Impact on Teaching Practice and Student Learning

From the perspective of the instructors, this project has introduced them to and afforded them an in-depth opportunity to explore a viable and at times more desirable alternative method for assessing STEM student work than traditional score-based grading. Standards-based grading made instructors think more explicitly about the connection between course work and course outcomes. The grading system appears to lie between cumulative score-based grading and more qualitative approaches such as mastery- or criterion-based grading in terms of instructor time and effort. The development of computer applications that directly support standards-based grading may be of assistance in this area—the goal of such applications would be to automate the gathering and dissemination of instructor feedback and standards development reports, which were manually done or within the constraints of existing software for this project.

From the student perspective, the most prominent qualitative feedback received was that standards-based grading more closely corresponded to how performance is evaluated in industry and advanced studies than cumulative scoring. The specificity of areas of strength and improvement helped students have better focus on which knowledge or skills can use more time or effort.

Relevance to Teaching in our Field and Beyond

With continuing feedback from industry and research institutions that STEM graduates do not always possess the appropriate knowledge and skill set for work in their areas upon leaving the undergraduate level, approaches that have the potential to bridge this gap more effectively remain relevant and of interest. Our study of standards-based grading in STEM undergraduate courses aligns with this broader need. The exploration of STEM course assessment methods in general will continue to carry this degree of relevance until the students that STEM programs produce both demonstrably graduate with the requisite expertise and constitute a better match for industry, research, or other primary endeavors that they might enter after they receive their bachelors degree.

Context of the Relevant Literature

Although standards-based grading has gained popularity at the K–12 level (Sadler, 2005; Tomlinson & McTighe, 2006; Carpinelli et al., 2008; Scirfiny, 2008), there have been no studies to date that examined the implementation of this specific system for undergraduate STEM education. A few undergraduate STEM studies have considered the importance of how a grading system and related instructor feedback impact student learning (Sharp et al., 1997; Catalano & Catalano, 1999; Meier et al., 2000; Moskal et al., 2002). A body of work that has considered grading systems investigated the use of mastery-based grading. Mastery-based grading is a system that provides students with multiple opportunities to demonstrate appropriate mastery of course concepts (Armacost & Pet-Armacost, 2003; Potter, R. & Meisels, G., 2005). Student improvements led to better grades, but the system still ultimately relies on
traditional, summative score-based grading. Alternatively, criterion-based assessment makes use of tasks that appraise student work using multiple, often interlocking criteria, which refer to qualities, characteristics, or attributes of a student response (Sadler, 1987). This approach addresses the need to provide student feedback and assesses students non-traditionally. Anomalous patterns arising with criterion-based assessment have been identified that suggest the overall process lacks robustness (Sadler, 2009). Standards-based grading accounts for the deficiencies of traditional score-based and criterion-based grading systems by providing meaningful feedback to students of their development toward a given objectives and standards.
Bibliography


Grigsby, M., 2009, College Life through the Eyes of Students, State University of New York Press, Albany, NY.

Huba, M.E., & Freed, J.E., 2000, Learner-centered assessment on college campuses: Shifting the focus from teaching to learning, Allyn & Bacon, Needham Heights, MA.


National Survey of Student Engagement (NSEE), 2007, Experiences that Matter: Enhancing Student Learning and Success, Center for Postsecondary Research, Indiana University, Bloomington, IN.


Suskie, L., 2004, Assessing student learning: A common sense approach, Anker, Boston, MA.

Tomlinson, C., & McTighe, J., 2006, Integrating differentiated instruction and understanding by design, Association for Supervision and Curriculum Development (ASCD), Alexandria, VA.


Dissemination
