A Collaborative Effort for Improving Calculus Through Better Assessment Practices

Justin Heavilin Kyle Hodson Brynja Kohler

December 18, 2015

Abstract

Like many institutions across the country, Utah State University's Department of Mathematics and Statistics has embarked on an effort to improve the calculus sequence with the following objectives: (1) improve our students' comprehension and application of key topics, (2) retain/recruit more students into STEM majors, and (3) provide more consistency across sections. After initial planning and preparation in the 2014-15 academic year, new practices were ready for implementation. In the fall of 2015, teams of instructors worked from common guided course notes, and met weekly to discuss instruction and develop common assessments. This poster displays the methodology of test design and item analysis we employed in the Calculus 2 course. While our team is only at the beginning stages of this work, the methods for creating reliable and relevant measures of student learning hold promise for achieving the goals of our reform.

1 Introduction

For many university students calculus courses serve as the gate-keeper that either attracts or repels them from pursuing STEM majors and professions, and is consequently a priority area for reform [1]. Here we share the background and motivations of our current work at Utah State University:

- The College of Engineering presented our department with data showing wildly inconsistent DWF rates in various sections of calculus by instructor.
- A Calculus Committee was formed and after a series of organizational meetings, we set about organizing the presentation, content, and evaluation in all courses and sections of Calculus.
- To provide cohesion between the sections of Calculus, we produced a set of guided lecture notes. These published notes act as a scaffold upon which instructors can build their lectures, as well as a learning tool for students during class sessions.
- To add consistency across all sections we scheduled common midterms and final exams. Exam questions are proposed and vetted by all instructors, and grading of exams was accomplished by assigning one problem to each faculty member thus ensuring consistent assessment of each test item.
- During weekly meetings instructors address questions about content and scope of material, propose exam questions, and share instructional techniques. These meetings also identified much-needed corrections to be included in the next iteration.
- In Calculus 1 and Calculus 2, we began the semester with a skills pre-test to help ascertain the level of preparedness of the students with varying backgrounds.

In this poster, we share our methods and findings as they pertain to the skills pre-test we offered at the beginning of the semester in Calculus 2. We also conducted item analysis after a midterm later in the semester and found improvement of test validity. Collaborative work throughout the semester was likely a quality professional development experience for all of the instructors involved.

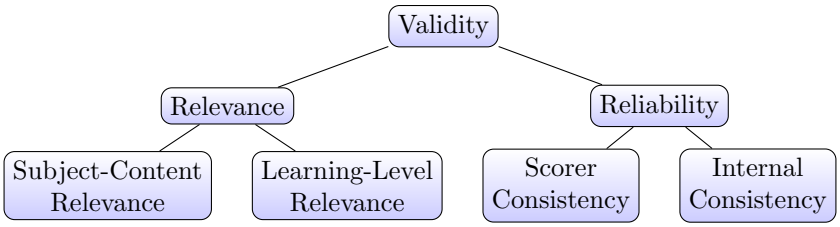

Figure 1: Cangelosi's hierarchy of test validity, dependent on relevance and reliability of test items.[3]

2 Methods

If we wish to maximize student achievement we then must pay greater attention to the improvement of classroom assessment. Instructors need training in creating better assessment tools. [2] To this end we established the following practices and methods.

- A common exam schedule was determined for all sections of Calculus 2, allowing for one exam to be administered to all six sections of the course.
- We agreed upon specific learning objectives to be tested. Instructors listed topics, but we also discussed the level of learning we were after and used Cangelosi's cognition scheme to guide our work. [3]
- Once we laid out a test blueprint of topics and their relative emphasis, committee members took ownership of test items for each topic. This entailed creating the test item, writing a scoring rubric, and then grading the item on all exams across all sections. This effort made scoring quick and consistent.
- Exam scores across all sections were collated for analysis (via boxplots, ANOVAs, and Tukey's HSD).
- Scoring of individual items as well as total scores permitted test item analysis. In addition to quantifying item difficulty and discrimination, we also calculuted Hoffman's Efficiency Coefficient for each item. Finally we compute the exam's Kuder–Richardson Formula 20 (KR-20) coefficient as a measure of internal reliability [3]. These provide guidelines for reliability of test questions, feeding back to future test design. Our poster will include the formulae and interpretations to illustrate this process for others.

We have not only seen our test reliability increase over the semester, but our level of collaboration overall has served as professional developing experience for both the graduate students and faculty. For example, as the semester progressed, faculty began observing one another's classes in an effort to better understand how material was presented by peers, and teachers shared supplemental materials to enhance lectures with computer animations and interactive learning activities.

3 Recommendations for the Next Cycle

As the world and calculus evolves, effective instruction will always require changes. This work is an iterative process, and should never be considered complete.

Extending beyond the traditional efforts to make common exams, we have made considerable strides in applying item analysis toward informing our exam design. While attention has been paid to instructional objectives and materials treatment of quantitative thinking skills, classroom assessments often fail to match these aspirations. Some students use tests to understand the teachers' expectations and prioritize topics. Thus poor quality assessments that fail to consider higher-order thinking skills will inhibit the development of those skills. [5] The item analysis provides a powerful tool for quantifying the validity of these assessments. Our work has has beneficial side effects such as motivating the empowering practice of observing each other's classes.

References

[1] Bressoud, D., Vilma M., & Rasmussen, C. 2015. *Insights and Recommendations from the MAA National Study of College Calculus*. MAA Notes. Washington, DC: The Mathematical Association of America.

[2] Stiggins, R. J. 2002. Assessment Crisis: The Absence of Assessment FOR Learning. *Kappan Professional Journal*.

[3] Cangelosi, J. 2003. *Teaching Mathematics in Secondary and Middle School: An Interactive Approach*. Pearson Education Inc. Upper Saddle River, NJ.

[4] Cangelosi J.S. 2000. *Assessment Strategies for Monitoring Student Achievement*. New York. Addison Wesley.

[5] Stiggins, R. J. Revitalizing Classroom Assessment: The Highest Priority. *Phi Delta Kappa*, 69, 363-368 1988.