A Comprehensive Assessment Program — Three Years Later

Kenneth B. Hannsgen
Department of Mathematics
Virginia Tech
Blacksburg, VA
hannsgen@vt.edu

Abstract. Recent progress is summarized in a departmental program of data collection and analysis at Virginia Tech. The assessment program and this report focus on service/general education courses in mathematics. Sketches of particular studies illustrate the developing awareness of the possibilities and limitations of assessment. Impact on decision-making in this large mathematics department is emphasized.

Background and Goals

Virginia Tech is a land grant, state, Type I research university, with an overall enrollment of over 25,000 students. The semester enrollment in courses offered by the mathematics department averages 8,000–12,000 students. About 90% of these students are registered in four 1000 to 2000 level calculus sequences designed to meet general education requirements as well as the specialized needs of students in i) math, physical sciences and engineering, ii) business, iii) life sciences, and iv) architecture.

Olin and Scruggs [1][1] reported on the comprehensive assessment program that Olin, as department head, had initiated in 1995. They discussed the ways in which a comprehensive assessment program helps the department to meet its responsibilities and document its achievements. They described collection and analysis of data to measure academic success across the service course spectrum, evaluation of several innovative programs, and monitoring of grading equity across course sections. Outcomes through 1997 were reported. We focus on recent developments in several of these areas.

In contrast to a carefully planned and controlled pilot study, this assessment program is an attempt at department-wide accounting. It was instituted by the department head and has been carried out "on the fly" by faculty members who are not experts and who have other major commitments and interests. To the extent that this situation is characteristic of what might take place in a large, research-oriented department, we hope that this report will help planners who embark on similar efforts. We emphasize the impact of the program on the department rather than the detailed content of any particular study.

Description and Findings

Data management. The process of data collection and analysis described by Olin and Scruggs has survived one change of department head and two changes of assessment coordinator. Mr. Kevin Bradley, a graduate student in the Department of Psychology, completed a two-year term in the latter job in 2002. A four-person faculty committee, chaired by the author, oversaw his work. We collected data on courses, students, tests, and grades. One of Bradley's main accomplishments was to organize all of this into a comprehensive, easily accessed 10-year database, to which we add new data each semester.

---

[1] www.maa. org/saum/maanotes49/224.html.

Common final exams. For each large multisection course, a three-person committee, consisting of recent but not current teachers of that course, writes a one-hour test that makes up half of the final exam. Considerations of grading effort and data management restrict us to multiple-choice format for assessment on this scale. Course coordinators evaluate the results and write short reports on student success in relation to our lists of course goals, and on the quality of the exam questions. They suggest improvements for the test and the course. This is the closest we normally come to a complete assessment cycle. The effort requires logistical coordination, as well as the participation of dozens of faculty, many of whom are uninterested in or skeptical about assessment.

A review of the physical science calculus sequence, conducted in 2002, confirmed that the system does not yet provide useful formative assessment. Different committees write the tests each semester, with the focus on writing an accurate final exam, useful for grading purposes. Fitting the test to the needs of long-term assessment is a secondary consideration at best, and exam writers rarely consult the historical record of previous exams and reports. Beyond this, our original goal lists have turned out to be too vague, and the connection between particular questions and particular goals is ambiguous at best. These problems also reflect the difficulty of finding agreement across the faculty on what the learning goals are and on whether particular test results demonstrate that they have been achieved.

We will next attempt, on a pilot basis, to reformulate the learning goals in a way that is more closely tied to specific test questions. (A natural pilot course is first-year integral calculus for the physical sciences, where we have some experience with competency quizzes.) For example, a goal concerning integration might indicate a difficulty level and specify definite integration rather than antiderivatives. A goal on Simpson's rule would specify whether a table of values or a formula is given. Tests will include a core of questions from a standard pool, so that test construction will be more centralized. There is a danger that we will overemphasize mechanical skills, since the variety of possible questions grows rapidly as one goes beyond these skills. Still, there is a consensus in the department that we need some standard barometer of student achievement.

Academic measurement and technology. Teaching innovations involving technology, including online instruction in a few courses, raise serious and controversial questions about learning achievement. The assessment program has made important contributions here.

The first course in the life sciences sequence (1015, precalculus) is given entirely online, but at first we included one weekly live class session to help orient students to the course and discuss applications. Was this live section worth its cost? In Fall 2000, students with relatively strong math skills, as indicated by SAT scores and high school grades, were given the option of taking the course with and without the live session. We compared grades in the course and downstream in the succeeding courses. The placement method had been designed with pedagogical aims rather than as a randomized trial for measurement purposes. Bradley's study used statistical means to compensate in part for these sampling problems. While the live-session students showed a slight advantage in final grades (attributable largely to differences in homework scores), there was no difference between live-session and "independent" students in final exam scores. Moreover, a follow-up study of students who went on to the next course (1016, differential calculus) showed no difference in pass rates between students from the two 1015 groups. Among the few students who crossed over from 1015 to differential calculus in the physical science/engineering sequence, those from the independent 1015 group actually did better. Since the main purpose of 1015 is preparation for subsequent courses, these results gave convincing evidence that the live sessions were not cost effective.

In 1997, the two-credit first-year course on linear algebra[2] shifted to an all-online mode for the entire 2000-student annual enrollment, excluding honors sections, using locally written software. The change yielded dramatic cost reductions, enabling the department to shift resources to other areas. (Reports on implementation, learning outcomes, and cost savings appear on the website of Pew Program in Course Redesign,[3] which supported the development project. A more detailed appendix on assessment is available on request from the author.) Our assessment statistics documented increases in the percentages of students achieving at least a C or at least a D– in the course. General grade levels remained steady through the change. Final exam results indicated that the topical area of eigenvalues and eigenvectors needed more attention, but no clear trend emerged in overall scores. This was not a controlled study, and the assessment results are open to varying interpretations; in particular, the syllabus and tests changed as course developers learned what they could do with the online medium. On balance, however, these results showed that expanded online efforts offer a reasonable way for the department to control costs and maintain effectiveness in lower-level courses, subject to the availability of startup

[2] course-delivery.emporium.math.vt.edu/courses/math1114/index.html

[3] www.center.rpi.edu/PewGrant/RD1award/VA.html

funding. In 2002–03 a second course of the life sciences sequence, differential calculus, went completely online in this way. In addition, the sequences for business and physical sciences now include some use of the online utilities.

The linear algebra course has also provided feedback on student attitudes, through an online survey that we conduct near the end of each semester. This is a long survey, with participation rates of over 90%. Questions address students' work habits and their satisfaction with various learning resources. For example, the fraction of students who agreed that the online tutorials "explained concepts well" increased from 71% to 88% over the years. The surveys help the department to track the performance of tutoring staff at the Math Emporium,[4] a large learning center where students do much of their work and testing in the course.

Special calculus sections. As reported in [1], Virginia Tech instituted an "emerging scholars" program (ESP) in 1996, modeled loosely on successful programs at other institutions, such as the University of California Davis.[5] Early success led to an expansion of the program to cover nearly all students predicted to be at risk in first-year engineering calculus. (In contrast to the Davis program and others, then, Virginia Tech Emerging Scholars was not restricted to a small, highly motivated subset of the student population.) As pilot funding from the university administration began to run out, the department turned to assessment results in order to decide whether to press for continuation of the program. Because of year-to-year changes and other factors, the data from this program were very messy, and Bradley's study went through numerous revisions as he and the assessment committee wrestled with questions and assumptions. The final report uses a variety of statistical stratagems to sort out the picture, and there are conflicting outcomes in several cases. Overall, attending the additional ESP hours helped the at-risk students to survive the first-year courses, but there was no carryover benefit (and perhaps even some negative effect) in the second-year courses (differential equations and multivariable calculus). For example, out of n = 2003 pairs of ESP and non-ESP students enrolled in the course between 1993 and 2001, matched on the basis of SAT scores and high school grades, 69% of the ESP students versus 60% of the non-ESP students successfully completed the first year of calculus, but only 27% of the ESP students versus 32% of the non-ESP students successfully completed the second year; these trends were confirmed in several other outcome measures. These results supported a decision not to commit further resources to the program in this form.

Insights

The experience at Virginia Tech is an example of learning the hard way. As traditional mathematics faculty, we are concerned and careful about evaluation of students, but systematic educational assessment is an unfamiliar and in some ways uncomfortable experience. We began with a new department head's need to understand what was happening in our classes (many of which involved experimentation) and to demonstrate the value of what we do to internal and external providers of funding for major changes. We collected data and compiled results, and the process began to yield some insights into the successes and challenges in our program. The availability of an organized body of data has enabled the department to respond to subsequent requests for assessment.

We have found that the numbers do not provide simple, conclusive answers. Results can shed light on outcomes and help in making decisions, but it is rare to find a smoking gun that resolves controversies or overcomes strongly held beliefs about pedagogy. We have gotten some useful large-scale information, but we have as yet found little information that suggests how we need to shift emphasis in the syllabus of any one class. We continue to seek better ways to incorporate assessment into the improvement process.

Acknowledgement. The author thanks Professor Peter Haskell for extensive comments and suggestions.

Reference

1. Olin, R. and Scruggs, L. "A comprehensive, proactive assessment program," in Gold, B., Keith, S. Z., and Marion, W. A. eds. *Assessment Practices in Undergraduate Mathematics,* MAA Notes #49. Washington, DC: Mathematical Association of America, 1999.

[4] www.emporium.vt.edu/

[5] www.math.ucdavis.edu/~kouba/KoubaHomepageDIRECTORY/ESP.html