Commentary

Assessment and Quality Control in Chemistry Education

by Thomas Holme

Increasingly, educators are facing demands for greater accountability. While some may chafe at the business analogy, accountability is essentially a move towards quality control in education. How do we know our efforts provide the level of quality that is demanded by our constituents? When students or parents ask us to establish that we have measured their knowledge with accuracy, how do we provide evidence? There are many strategies for assessment and it is appropriate to argue that we need more than a snapshot of knowledge provided by a single exam. Nonetheless, for decades the Division of Chemical Education has provided a reliable means of testing knowledge through the tests constructed by the Examinations Institute. How are these exams constructed? How can they fit into an assessment program that meets current demands for accountability?

The Exam Development Process

One of the more exigent variables associated with knowledge assessment is the definition of what is important to assess. For example, who decides what portion of a course that involves 30 or more weeks of instruction is important enough to test? This question points to the importance of strategies that involve regular measurement of student learning, but it also identifies how daunting it is to develop a meaningful final exam—especially if you are doing so on your own. Is your final exam a good measure of student learning in your course? If you had to defend it, how would you do so?

Identifying and Writing Questions

Exams produced by the Exams Institute are constructed by committees of volunteers from around the U.S. Committee members regularly teach the course for which the exam is intended. The initial meeting of a committee is largely spent debating the key initial question: What topics in this course merit inclusion on the exam? Because the choice is made by a group whose individual members often argue vigorously either for or against a particular subject, the content coverage of ACS DivCHED exams inevitably reflects national trends in instruction of a course.

This writing process also speaks directly to the issue of quality. If a group of peers has, by debate, arrived at a set of questions to reflect the content coverage of a course, there is some assurance of quality in education when students perform well on this exam. A student who scores above average on an exam from the Institute can legitimately claim to have met or exceeded a reasoned standard of learning that has been verified in a manner that is external to the specific environment of the course that was taken. At least from a content perspective, the exam has been set within a rather broad consensus of what is important for the student to know upon completion of that course.

Once a set of topics has been agreed upon, committee members work individually to write possible multiple-choice questions for the exam being developed. Writing good multiple-choice questions is challenging, but the advantages of fast and accurate grading render them an appropriate choice for exams meant to be used for comparisons among students nationally.

Honing and Selecting Questions

There are two critical steps that provide the kind of quality assurance needed for an examination to be valid in a national sampling of students. First, questions are massaged by the committees after they have been composed by the individual members. For many of our volunteers, the process of improving a question—identifying unexpected concepts that were included in the original wording, choosing distractors, and editing the wording for concise and clear statements—makes committee service a significant learning experience. Virtually all Exams Institute committee volunteers report increased confidence in their test writing ability for their own courses after they complete their committee service.

Second, the committee writes substantially more than twice the number of questions that will appear in the final version of the exam. They ultimately set two trial versions of the exam, which are tested by volunteers (from the committee and elsewhere) in their classes. With this method twice the number of questions that are needed on the final published version of the exam are tested, and statistical analysis helps reveal strengths and weaknesses in the questions that were set. Thus, an item on the published exam has not only been improved by the deliberations of the committee members, it has also been tested for its validity among national samples of students.

Validity

While the limitations of multiple-choice formats are noteworthy, the process used to derive the specific items on Exams Institute products is a robust one. Quite importantly, the process provides quality assurance to any instructor who elects to use the exam because several levels of validity have been built in. The content validity is assured by the deliberations of the committee of volunteers. The measurement validity is assured by the nature of the statistical analysis to which each question has been subjected. Finally, the overall test validity—the combination of items that composes the overall exam—is built by the compilation of statistics from users around the country that results in the production of norms. Once published, these norms allow one form of external verification of the learning that has taken place in a course, and thus the tests reflect on the type of quality that students, parents, and administrators demand of instructors these days.

Exams and Security

Despite the various procedures that are designed to assure the validity of Institute exams, there are threats to that validity. Perhaps the most apparent one is security. If students gain access to items prior to taking the exam, the claims of validity are compromised. For this reason, the Exams Institute takes significant precautions to maintain security and must expect the same from its users.

Recently, for example, an instructor in Texas was approached by his students who had noted the Organic Exam they were going to take had been posted on the Web. The site of the posting was an East Coast university, but the Internet crosses borders and the actions of one faculty member led to the compromise of this exam for all. Exams Institute staff regularly search the Internet for possible posting of secure materials, but had missed this site. Had the students not informed their instructor but instead had taken the exam for which they knew the answers, the test's validity as an assessment of knowledge would have been nonexistent.[1]

This is an egregious example of the dangers of compromising the security of an exam. There are other less obvious threats as well. Exam instructions, as printed on the cover of the test booklet, call for no programmable calculators. Many instructors expect that this requirement is designed to make sure students do not load equations or other information into their calculators to gain an advantage while taking the exam. This is a legitimate concern for instructors, but from the perspective of the Exams Institute, we are equally concerned that students do not program questions into their calculators and take them *out* of the exam room. If a concerted effort were made by a group of students, such a practice could lead to the compromise of an exam—perhaps only locally within a single school or region—but in a way that would invalidate the kind of quality assurance one hopes to gain by using a standardized exam. For the same reason, recent exams also forbid the carrying of PDA devices into an exam—mostly because of the information that could be carried out.

For many students the mystery associated with a standardized exam adds to their anxiety. From a human factors' perspective, anxiety presents an extra variable that affects the measurement of knowledge in unpredictable ways. It is, however, reasonable to suppose that the anxiety increase associated with taking a standardized exam in a course falls along some continuum that is reflected in the sample of students whose scores were used to set the national norms. To some extent, therefore, the anxiety factor (and other random variables associated with knowledge testing) is built into the measurement instrument when comparisons are made.

Uses and Abuses of Exams

The clamor for quality control (accountability) not only illuminates some of the opportunities for the use of Exams Institute products, it also provides for the real possibility of abuses. There is an inclination in some sectors of the public to say that student performance alone is the measure of effective teaching. Long time users of Exams Institute tests note fairly wide variation in performance by different classes that are difficult to attribute solely to the instructor—for example, the same course in different years. As such, factoring in the performance of an individual class relative to national norms in a merit-based performance review (such as tenure or promotion) is problematic. At its best, the exam provides a calibrated snapshot of student performance, and such an item might help an instructor improve his or her teaching, but it would be difficult to use as hard evidence for inadequacies given the variability of student populations from year to year or from section to section. I have argued here that the quality of our calibration of this snapshot of student knowledge is robust, but this does not argue that exams from the Institute ought to be arbiters of teaching excellence. They provide a piece of evidence, a component of a broader portfolio, that can be used to assure parents and students about the quality of their education.

What, then, are the options for an instructor who wishes to include exams from the Institute in an overall scheme to measure quality education? In fact they are many. Consider the rather natural inclination to note that the exam doesn't test everything an instructor might want to include. Some feel that they cannot get an adequate feel for their students' understanding without long-answer questions. At least at the General Chemistry level, it may be possible to do both. The Exams Institute constructs a "brief" version of the full-year general chemistry exam. The initial motivation for this exam was for it to be available in locations where final exams are held during class periods. If, however, a two-hour time slot is available, the brief exam would allow an instructor to calibrate against national norms (taking half the time period) and then leave the second half of the exam period for more individualized questions that assess specific interests of the instructor, perhaps in long-answer format.

Assessment of teaching innovation presents another arena where our exams can provide important calibration. These exams are not designed to measure the specifics of a particular teaching innovation, but they can be used to show that the new methods "do no harm". Establishing the benefits of an innovation is often a very challenging task, but while that type of assessment process is being established and verified, it may be worthwhile for an innovator to use ACS Exams to show that student performance remains at or above levels comparable to national norms. Again, the nationally normed exams provide evidence of quality, a feature of particular importance when trying to establish the utility of an educational reform or innovation.

On the Horizon

There are also new horizons facing the Exams Institute, and we hope to provide even more tools for assessment in the future. We are embracing the challenge of creating new electronically delivered assessment tools that maintain both the quality that we derive from the committee-based process described above and the security that maintains that quality. We are always looking for interested volunteers to help us move forward with new exams and new products. We look

Commentary

forward to hearing from dedicated educators who might be interested in helping move the Exams Institute forward. Send us an email at *chmexams@uwm.edu* and let us know how you might want to get involved.

Note

1. Users of the Organic Exams will have received information about a replacement program for the exams that were compromised by this incident.

Thomas Holme is a member of the Department of Chemistry, University of Wisconsin–Milwaukee, Milwaukee, WI 53201; tholme@uwm.edu. *He is director of the DivCHED Examinations Institute,* chmexams@uwm.edu; http://www.uwm.edu/Dept/chemexams/.