Asking the Right Questions

Lynn Arthur Steen
Department of Mathematics, Statistics, and Computer Science
St. Olaf College
steen@stolaf.edu

Assessment is about asking and answering questions. For students, "how am I doing?" is the focus of so-called "formative" assessment, while "what's my grade?" often seems to be the only goal of "summative" assessment. For faculty, "how's it going?" is the hallmark of within-course assessment using instruments such as ten-minute quizzes or one-minute responses on 3×5 cards at the end of each class period. Departments, administrations, trustees, and legislators typically ask questions about more aggregated levels: they want to know not about individual students but about courses, programs, departments, and entire institutions.

The conduct of an assessment depends importantly on who does the asking and who does the answering. Faculty are accustomed to setting the questions and assessing answers in a context where outcomes count for something. When assessments are set by someone other than faculty, skepticism and resistance often follow. And when tests are administered for purposes that don't "count," (for example, sampling to assess general education or to compare different programs), student effort declines and results lose credibility.

The assessment industry devotes considerable effort to addressing a variety of similar contextual complications, such as:

- different purposes (diagnostic, formative, summative, evaluative, self-assessment, ranking);
- different audiences (students, teachers, parents, administrators, legislators, voters);
- different units of analysis (individual, class, subject, department, college, university, state, nation);
- different types of tests (multiple choice, open ended, comprehension, performance-based, timed or untimed, calculator permitted, individual or group, seen or unseen, external, written or oral);
- different means of scoring (norm-referenced, criterion referenced, standards-based, curriculum-based);
- different components (quizzes, exams, homework, journals, projects, presentations, class participation);
- different standards of quality (consistency, validity, reliability, alignment);
- different styles of research (hypothesis-driven, ethnographic, comparative, double-blind, epidemiological).

Distinguishing among these variables provides psychometricians with several lifetimes' agenda of study and research. All the while, these complexities cloud the relation of answers to questions and weaken inferences drawn from resulting analyses.

These complications notwithstanding, questions are the foundation on which assessment rests. The assessment cycle begins with and returns to goals and objectives

(CUPM, 1995). Translating goals into operational questions is the most important step in achieving goals since without asking the right questions we will never know how we are doing.

Two Examples

In recent years two examples of this truism have been in the headlines. The more visible—because it affects more people—is the new federal education law known as No Child Left Behind (NCLB). This law seeks to ensure that every child is receiving a sound basic education. With this goal, it requires assessment data to be disaggregated into dozens of different ethnic and economic categories instead of typical analyses that report only single averages. NCLB changes the question that school districts need to answer from "What is your average score?" to "What are the averages of every subgroup?" Theoretically, to achieve its titular purpose, this law would require districts to monitor every child according to federal standards. The legislated requirement of multiple subgroups is a political and statistical compromise between theory and reality. But even that much has stirred up passionate debate in communities across the land.

A related issue that concerns higher education has been simmering in Congress as it considers reauthorizing the law that, among other things, authorizes federal grants and loans for postsecondary education. In the past, in exchange for these grants and loans, Congress asked colleges and universities only to demonstrate that they were exercising proper stewardship of these funds. Postsecondary institutions and their accrediting agencies provided this assurance through financial audits to ensure lack of fraud and by keeping default rates on student loans to an acceptably low level.

But now Congress is beginning to ask a different question. If we give you money to educate students, they say, can you show us that you really are educating your students? This is a new question for Congress to ask, although it is one that deans, presidents, and trustees should ask all the time. The complexities of assessment immediately jump to the foreground. How do you measure the educational outcomes of a college education? As important, what kinds of assessments would work effectively and fairly for all of the 6,600 very different kinds of postsecondary institutions in the United States, ranging from 200-student beautician schools to 40,000-student research universities? Indicators most often discussed include the rates at which students complete their degrees or the rates at which graduates secure professional licensing or certification. In sharp contrast, higher education mythology still embraces James Garfield's celebrated view of education as a student on one end of a log with Mark Hopkins on the other end. In today's climate of public accountability, colleges and universities need to "make peace" with citizens' demand for candor and openness anchored in data (Ekman, 2004).

I cite these examples to make two points. First, the ivory tower no longer shelters education from external demands for accountability. Whether faculty like it or not, the public is coming to expect of education the same kind of transparency that it is also beginning to demand of government and big business. Especially when public money is involved—as it is in virtually every educational institution—public questions will follow.

Second, questions posed by those outside academe are often different from those posed by educators, and often quite refreshing. After all these years in which school districts reported and compared test score averages, someone in power finally said "but what about the variance?" Are those at the bottom within striking distance of the average, or are they hopelessly behind with marks cancelled out by accelerated students at the top? And after all these years of collecting tuition and giving grades, someone in power has finally asked colleges and universities whether students are receiving the education they and the public paid for. Asking the right questions can be a powerful lever for change, and a real challenge to assessment.

Mathematics

One can argue that mathematics is the discipline most in need of being asked the right new questions. At least until very recently, in comparison with other school subjects mathematics has changed least in curriculum, pedagogy, and assessment. The core of the curriculum in grades 10–14 is a century-old enterprise centered on algebra and calculus, embroidered with some old geometry and new statistics. Recently, calculus passed through the gauntlet of reform and emerged only slightly refurbished. Algebra—at least that part known incongruously as "College Algebra"—is now in line for its turn at the reform carwash. Statistics is rapidly gaining a presence in the lineup of courses taught in grades 10–14, although geometry appears to have lost a bit of the curricular status that was provided by Euclid for over two millennia.

When confronted with the need to develop an assessment plan, mathematics departments generally take this traditional curriculum for granted and focus instead on how to help students through it. However, when they ask for advice from other departments, mathematicians are often confronted with rather different questions (Ganter & Barker, 2004):

- *Do students in introductory mathematics courses learn a balanced sample of important mathematical tools?*
- *Do these students gain the kind of experience in modeling and communication skills needed to succeed in other disciplines?*
- *Do they develop the kind of balance between computational skills and conceptual understanding appropriate for their long-term needs?*
- *Why can't more mathematics problems employ units and realistic measurements that reflect typical contexts?*

These kinds of questions from mathematics' client disciplines strongly suggest the need for multi-disciplinary participation in mathematics departments' assessment activities.

Similar issues arise in relation to pedagogy, although here the momentum of various "reform" movements of the last two decades (in using technology, in teaching calculus, in setting K–12 standards) has energized considerable change in mathematics instruction. Although lectures, problem sets, hour tests, and final exams remain the norm for mathematics teaching, innovations involving calculators, computer packages, group projects, journals, and various mentoring systems have enriched the repertoire of postsecondary mathematical pedagogy. Many assessment projects seek to compare these new methods with traditional approaches. But client disciplines and others in higher education press even further:

- *Do students learn to use mathematics in interdisciplinary or "real-world" settings?*
- *Are students encouraged (better still, required) to engage mathematics actively in ways other than through routine problem sets?*
- *Do mathematics courses leave students feeling empowered, informed, and responsible for using mathematics as a tool in their lives?* (Ramaley, 2003)

Prodded by persistent questions, mathematicians have begun to think afresh about content and pedagogy. In assessment however, mathematics still seems firmly anchored in hoary traditions. More than most disciplines, mathematics is defined by its problems and examinations, many with histories that are decades or even centuries old. National and international mathematical Olympiads, the William Lowell Putnam undergraduate exam, the Cambridge University mathematics Tripos, not to mention popular problems sections in most mathematics education periodicals attest to the importance of problems in defining the subject and identifying its star pupils. The correlation is far from perfect: not every great mathematician is a great problemist, and many avid problemists are only average mathematicians. Some, indeed, are amateurs for whom problem solving is their only link to a past school love. Nonetheless, for virtually everyone associated with mathe-

Greece, 250 BCE

If thou art diligent and wise, O stranger, compute the number of cattle of the Sun, who once upon a time grazed on the fields of the Thrinacian isle of Sicily, divided into four herds of different colours, one milk white, another a glossy black, a third yellow and the last dappled. In each herd were bulls, mighty in number according to these proportions: Understand, stranger, that the white bulls were equal to a half and a third of the black together with the whole of the yellow, while the black were equal to the fourth part of the dappled and a fifth, together with, once more, the whole of the yellow. Observe further that the remaining bulls, the dappled, were equal to a sixth part of the white and a seventh, together with all of the yellow. These were the proportions of the cows: The white were precisely equal to the third part and a fourth of the whole herd of the black; while the black were equal to the fourth part once more of the dappled and with it a fifth part, when all, including the bulls, went to pasture together. Now the dappled in four parts were equal in number to a fifth part and a sixth of the yellow herd. Finally the yellow were in number equal to a sixth part and a seventh of the white herd. If thou canst accurately tell, O stranger, the number of cattle of the Sun, giving separately the number of well-fed bulls and again the number of females according to each colour, thou wouldst not be called unskilled or ignorant of numbers, but not yet shalt thou be numbered among the wise.

But come, understand also all these conditions regarding the cattle of the Sun. When the white bulls mingled their number with the black, they stood firm, equal in depth and breadth, and the plains of Thrinacia, stretching far in all ways, were filled with their multitude. Again, when the yellow and the dappled bulls were gathered into one herd they stood in such a manner that their number, beginning from one, grew slowly greater till it completed a triangular figure, there being no bulls of other colours in their midst nor none of them lacking. If thou art able, O stranger, to find out all these things and gather them together in your mind, giving all the relations, thou shalt depart crowned with glory and knowing that thou hast been adjudged perfect in this species of wisdom.

—Archimedes. *Counting the Cattle of the Sun* matics education, assessing mathematics means asking students to solve problems.

Mathematical Problems

Problems on mathematics exams have distinctive characteristics that are found nowhere else in life. They are stated with precision intended to ensure unambiguous interpretation. Many are about abstract mathematical objects—numbers, equations, geometric figures—with no external context. Others provide archetype contexts that are not only artificial in setting (e.g., rowing boats across rivers) but often fraudulent in data (invented numbers, fantasy equations). In comparison with problems people encounter in their work and daily lives, most problems offered in mathematics class, like shadows in Plato's allegorical cave, convey the illusion but not the substance of reality.

Little has changed over the decades or centuries. Problems just like those of today's texts (only harder) appear in manuscripts from ancient Greece, India, and China (see sidebars). In looking at undergraduate mathematics exams from 100 or 150 years ago, one finds few surprises. Older exams typically include more physics than do exams of today, since in earlier years these curricula were closely linked. Mathematics course exams from the turn of the twentieth century required greater virtuosity in accurate lengthy calculations. They were, after all, set for only 5% of the population, not the 50% of today. But the central substance of the mathematics tested and the distinctive rhetorical nature of problems are no different from typical problems found in today's textbooks and mainstream exams.

Questions suitable for a mathematics exam are designed to be unambiguous, to have just one correct answer (which may consist of multiple parts), and to avoid irrelevant distractions such as confusing units or complicated numbers. Canonical problems contain enough information and not an iota more than what is needed to determine a solution. Typical tests are time-constrained and include few problems that students have not seen before; most tests have a high proportion of template problems whose types students have repeatedly practiced. Mathematician and assessment expert Ken Houston of the University of Ulster notes that these types of mathematics tests are a "rite of passage" for students around the world, a rite, he adds, that is "never to be performed again" once students leave university. Unfortunately, Houston writes, "learning mathematics for the principal purpose of passing examinations often leads to surface learning, to memory learning alone, to learning that can only see small parts and not the whole of a subject, to learning wherein many of the skills and much of the knowledge required to be a working mathematician are overlooked" (Houston, 2001).

All of which suggests a real need to assess mathematics assessment. Some issues are institutional:

- *Do institutions include mathematical or quantitative proficiency among their educational goals?*
- *Do institutions assess the mathematical proficiency of all students, or only of mathematics students?*

Others are more specifically mathematical:

- *Can mathematics tests assess the kinds of mathematical skills that society needs and values?*
- *What kinds of problems would best reflect the mathematical needs of the average educated citizen?*
- *Can mathematics faculty fairly assess the practice of mathematics in other disciplines? Should they?*

China, 100 CE

- A good runner can go 100 paces while a poor runner covers 60 paces. The poor runner has covered a distance of 100 paces before the good runner sets off in pursuit. How many paces does it take the good runner before he catches up to the poor runner?
- A cistern is filled through five canals. Open the first canal and the cistern fills in 1/3 day; with the second, it fills in 1 day; with the third, in 2 1/2 days; with the fourth, in 3 days, and with the fifth in 5 days. If all the canals are opened, how long will it take to fill the cistern?
- There is a square town of unknown dimensions. There is a gate in the middle of each side. Twenty paces outside the North Gate is a tree. If one leaves the town by the South Gate, walks 14 paces due south, then walks due west for 1775 paces, the tree will just come into view. What are the dimensions of the town?
- There are two piles, one containing 9 gold coins and the other 11 silver coins. The two piles of coins weigh the same. One coin is taken from each pile and put into the other. It is now found that the pile of mainly gold coins weighs 13 units less than the pile of mainly silver coins. Find the weight of a silver coin and of a gold coin.

— *Nine Chapters on the Mathematical Art*

India, 400 CE

- One person possesses seven asava horses, another nine haya horses, and another ten camels. Each gives two animals, one to each of the others. They are then equally well off. Find the price of each animal and the total value of the animals possessed by each person.
- Two page-boys are attendants of a king. For their services one gets 13/6 dinaras a day and the other 3/2. The first owes the second 10 dinaras. Calculate and tell me when they have equal amounts.

— *The Bakhshali Manuscript*

Issues and Impediments

Assessment has had a tenuous impact in higher education, especially among mathematicians who are trained to demand rigorous inferences that are rarely attainable in educational assessment. Some mathematicians are unrelentingly critical of any educational research that does not closely approach medicine's gold standard of randomized, double blind, controlled, hypothesis-driven studies. Their fears are not unwarranted. For example, a recent federal project aimed at identifying high quality educational studies found that only one of 70 studies of middle school mathematics curricula met the highest standards for evidence (*What Works*, 2005). Virtually all assessment studies undertaken by mathematics departments fall far short of mathematically rigorous standards and are beset by problems such as confounding factors and attrition. Evidence drawn entirely from common observational studies can never do more than suggest an hypothesis worth testing through some more rigorous means.

Notwithstanding skepticism from mathematicians, many colleges have invested heavily in assessment; some have even made it a core campus philosophy. In some cases this special focus has led these institutions to enhanced reputations and improved financial circumstances. Nonetheless, evidence of the relation between formal assessment programs and quality education is hard to find. Lists of colleges that are known for their commitment to formal assessment programs and those in demand for the quality of their undergraduate education are virtually disjoint.

Institutions and states that attempt to assess their own standards rigorously often discover large gaps between rhetoric and reality. Both in secondary and postsecondary education, many students fail to achieve the rhetorical demands of high standards. But since it is not politically or emotionally desirable to brand so many students as failures, institutions find ways to undermine or evade evidence from the assessments. For example, a recent study shows that on average, high stakes secondary school exit exams are pegged at the 8th and 9th grade level to avoid excessive failure rates (Achieve, 2004). Higher education typically solves its parallel problem either by not assessing major goals or by doing so in a way that is not a requirement for graduation.

- *How, if at all, are the mathematical, logical, and quantitative aspects of an institution's general education goals assessed?*
- *How can the goals of comprehending and communicating mathematics be assessed?*

When mathematicians and test experts do work together to develop meaningful assessment instruments, they confront major intellectual and technical hurdles. First are issues about the harmony of educational and public purposes:

- *Can a student's mathematical proficiency be fairly measured along a single dimension?*
- *What good is served by mapping a multifaceted profile of strengths and weaknesses into a single score?*

Clearly there are such goods, but they must not be oversold. They include facilitating the allocation of scarce educational resources, enhancing the alignment of graduates with careers, and —with care—providing data required to properly manage educational programs. They do not (and thus should not) include firm determination of a student's future educational or career choices. To guard against misuse, we need always to ask and answer:

- *Who benefits from the assessment?*
- *Who are the stakeholders?*
- *Who, indeed, owns mathematics?*

Mathematical performance embraces many different cognitive activities that are entirely independent of content. If content such as algebra and calculus represents the nouns—the "things" of mathematics—cognitive activities are the verbs: know, calculate, investigate, invent, strategize, critique, reason, prove, communicate, apply, generalize. This varied landscape of performance expectations opens many questions about the purpose and potential of mathematics examinations. For example:

- *Should mathematics exams assess primarily students' ability to perform procedures they have practiced or their ability to solve problems they have not seen before?*

- *Can ability to use mathematics in diverse and novel situations be inferred from mastery of template procedures?*
- *If learned procedures dominate conceptual reasoning on tests, is it mathematics or memory that is really being assessed?*

Reliability and Validity

A widely recognized genius of American higher education is its diversity of institutions: students' goals vary, institutional purposes vary, and performance standards vary. Mathematics, on the other hand, is widely recognized as universal; more than any other subject, its content, practices, and standards are the same everywhere. This contrast between institutional diversity and discipline universality triggers a variety of conflicts regarding assessment of undergraduate mathematics.

Assessment of school mathematics is somewhat different from the postsecondary situation. Partly because K–12 education is such a big enterprise and partly because it involves many legal issues, major assessments of K–12 education are subject to many layers of technical and scholarly review. Items are reviewed for, among other things, accuracy, consistency, reliability, and (lack of) bias. Exams are reviewed for balance, validity, and alignment with prescribed syllabi or standards. Scores are reviewed to align with expert expectations and desirable psychometric criteria. The results of regular assessments are themselves assessed to see if they are confirmed by subsequent student performance. Even a brief examination of the research arms of major test producers such as ETS, ACT, or McGraw Hill reveal that extensive analyses go into preparation of educational tests.

In contrast, college mathematics assessments typically reflect instructors' beliefs about subject priorities more than any external benchmarks or standards of quality. This difference in methodological care between major K-12 assessments and those that students encounter in higher education cannot be justified on the grounds of differences in the "stakes" for students. Sponsors of the SAT and AP exams take great pains to ensure quality control in part because the consequences of mistakes on students' academic careers are so great. The consequences for college students of unjustified placement procedures or unreliable final course exams are just as great.

- *Are "do-it-yourself" assessment instruments robust and reliable?*
- *Can externally written ("off the shelf") assessment instruments align appropriately with an institution's distinctive goals?*
- *Can locally written exams that have not been subjected to rigorous reviews for validity, reliability, and alignment produce results that are valid, reliable, and aligned with goals?*

Professional test developers go to considerable and circuitous lengths to score exams in a way that achieves certain desirable results. For example, by using a method known as "item response theory" they can arrange the region of scores with largest dispersion to surround the passing (so-called "cut") score. This minimizes the chance of mistaken actions based on passing or failing at the expense of decreased reliability, say, of the difference between B+ and A– (or its numerical equivalent).

- *How are standards of performance—grades, cut-scores—set?*
- *Is the process of setting scores clear and transparent to the test-takers?*
- *Is it reliable and valid?*

Without the procedural checks and balances of the commercial sector, undergraduate mathematics assessment is rather more like the Wild West—a libertarian free-for-all with few rules and no established standards of accountability. In most institutions, faculty just make up tests based on a mixture of experience and hunch, administer them without any of the careful reviewing that is required for development of commercial tests, and grade them by simply adding and subtracting arbitrarily assigned points. These points translate into grades (for courses) or enrollments (for placement exams) by methods that can most charitably be described as highly subjective.

Questions just pour out from any thoughtful analysis of test construction. Some are about the value of individual items:

- *Can multiple choice questions truly assess mathematical performance ability or only some correlate? Does it matter?*
- *Can open response tasks be assessed with reliability sufficient for high-stakes tests?*
- *Can problems be ordered consistently by difficulty?*
- *Is faculty judgment of problem difficulty consistent with empirical evidence from student performance?*
- *What can be learned from easy problems that are missed by good students?*

Others are about the nature and balance of tests that are used in important assessments:

- *Is the sampling of content on an exam truly representative of curricular goals?*
- *Is an exam well balanced between narrow items that focus on a single procedure or concept and broad items*

that cut across domains of mathematics and require integrated thinking?
- *Does an assessment measure primarily what is most important to know and be able to do, or just what is easiest to test?*

Interpreting test results

Public interest in educational assessment focuses on numbers and scores—percent passing, percent proficient, percent graduating. Often dismissed by educators as an irrelevant "horse race," public numbers that profile educational accomplishment shape attitudes and, ultimately, financial support. K–12 is the major focus of public attention, but as we have noted, pressure to document the performance of higher education is rising rapidly.

Testing expert Gerald Bracey warns about common misinterpretations of test scores, misinterpretations to which politicians and members of the public are highly susceptible (Bracey, 2004). One arises in comparative studies of different programs. Not infrequently, results from classes of different size are averaged to make overall comparisons. In such cases, differences between approaches may be entirely artificial, being merely artifacts created by averaging classes of different sizes.

Comparisons are commonly made using the rank order of students on an assessment (for example, the proportion from a trial program who achieve a proficient level). However, if many students are bunched closely together, ranks can significantly magnify slight differences. Comparisons of this sort can truly make a mountain out of a molehill.

Another of Bracey's cautions is of primary importance for K–12 assessment, but worth noting here since higher education professionals play a big role in developing and assessing K–12 mathematics curricula. It is also a topic subject to frequent distortion in political contests. The issue is the interpretation of nationally normed tests that report percentages of students who read or calculate "at grade level." Since grade level is defined to be the median of the group used to norm the test, an average class (or school) will have half of its students functioning below grade level and half above. It follows that if 30% of a school's eighth grade students are below grade level on a state mathematics assessment, contrary to frequent newspaper innuendos, that may be a reason for cheer, not despair.

Bracey's observations extend readily to higher education as well as to other aspects of assessment. They point to yet more important questions:

- *To what degree should results of program assessments be made public?*
- *Is the reporting of results appropriate to the unit of analysis (student, course, department, college, state)?*
- *Are the consequences attached to different levels of performance appropriate to the significance of the assessment?*

Program Assessment

As assessment of student performance should align with course goals, so assessment of programs and departments should align with program goals. But just as mathematics' deep attachment to traditional problems and traditional tests often undermines effective assessment of contemporary performance goals, so departments' unwitting attachment to traditional curriculum goals may undermine the potential benefits of thorough, "gloves off" assessment. Asking "how can we improve what we have been doing?" is better than not asking at all, but all too often this typical question masks an assumed *status quo* for goals and objectives. Useful assessment needs to begin by asking questions about goals.

Many relevant questions can be inferred from *Curriculum Guide 2004,* a report prepared recently by MAA's Committee on the Undergraduate Program in Mathematics (CUPM, 2005). Some questions—the first and most important—are about students:

- *What are the aspirations of students enrolled in mathematics courses?*
- *Are the right students enrolled in mathematics, and in the appropriate courses?*
- *What is the profile of mathematical preparation of students in mathematics courses?*

Others are about placement, advising, and support:

- *Are students taking the best kind of mathematics to support their career goals?*
- *Are students who do not enroll in mathematics doing so for appropriate reasons?*

Still others are about curriculum:

- *Do program offerings reveal the breadth and interconnections of the mathematical sciences?*
- *Do introductory mathematics courses contain tools and concepts that are important for all students' intended majors?*
- *Can students who complete mathematics courses use what they have learned effectively in other subjects?*
- *Do students learn to comprehend mathematically-rich texts and to communicate clearly both in writing and orally?*

A consistent focus of this report and its companion "voices of partner disciplines" (mentioned above) is that the increased spread of mathematical methods to fields well beyond physics and engineering requires that mathematics departments promote interdisciplinary cooperation both for faculty and students. Mathematics is far from the only discipline that relies on mathematical thinking and logical reasoning.

- *How is mathematics used by other departments?*
- *Are students learning how to use mathematics in other subjects?*
- *Do students recognize similar mathematical concepts and methods in different contexts?*

Creating a Culture of Assessment

Rarely does one find faculty begging administrators to support assessment programs. For all the reasons cited above, and more, faculty generally believe in their own judgments more than in the results of external exams or structured assessments. So the process by which assessment takes root on campus is more often more top down than bottom up.

A culture of assessment appears to grow in stages (North Central Assoc., 2002). First is an articulated commitment involving an intention that is accepted by both administrators and faculty. This is followed by a period of mutual exploration by faculty, students, and administration. Only then can institutional support emerge conveying both resources (financial and human) and structural changes necessary to make assessment routine and automatic. Last should come change brought about by insights gleaned from the assessment. And then the cycle begins anew.

Faculty who become engaged in this process can readily interpret their work as part of what Ernest Boyer called the "scholarship of teaching," (Boyer, 1990) thereby avoiding the fate of what Lee Shulman recently described as "drive-by teachers" (Shulman, 2004). Soon they are asking some troubling questions:

- *Do goals for student learning take into account legitimate differences in educational objectives ?*
- *Do faculty take responsibility for the quality of students' learning?*
- *Is assessment being used for improvement or only for judgment?*

Notwithstanding numerous impediments, assessment is becoming a mainstream part of higher education programs, scholarship, and literature. In collegiate mathematics, however, assessment is still a minority culture beset by ignorance, prejudice, and the power of a dominant discipline backed by centuries of tradition. Posing good questions is an effective response, especially to mathematicians who pride themselves on their ability to solve problems. The key to convincing mathematicians that assessment is worthwhile is not to show that it has all the answers but that it is capable of asking the right questions.

References

Achieve, Inc. (2004). *Do Graduation Tests Measure Up*? Washington, DC.

"Assessment of Student Academic Achievement: Levels of Implementation." (2002). *Handbook of Accreditation*. Chicago, IL: North Central Association Higher Learning Commission.

Boyer, Ernest L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Bracey, Gerald W. (2004). "Some Common Errors in Interpreting Test Scores." *NCTM News Bulletin*, April, p. 9.

Committee on the Undergraduate Program in Mathematics. (1995). "Assessment of Student Learning for Improving the Undergraduate Major in Mathematics," Focus, 15:3 (June) 24-28. www.maa.org/ saum/maanotes49/279.html

CUPM Curriculum Guide 2004. (2004). Washington, DC: Mathematical Association of America.

Ekman, Richard. (2004). "Fear of Data." *University Business*, May, p. 104.

Ganter, Susan and William Barker, eds. (2004). *Curriculum Foundations Project: Voices of the Partner Disciplines*. Washington, DC: Mathematical Association of America.

Houston, Ken. (2001). "Assessing Undergraduate Mathematics Students." In *The Teaching and Learning of Mathematics at University Level: An ICMI Study*. Derek Holton (Ed.). Dordrecht: Kluwer Academic, p. 407–422.

Ramaley, Judith. (2003). *Greater Expectations*. Washington, DC: Association of American Colleges and Universities.

Shulman, Lee S. (2004). "A Different Way to Think About Accountability: No Drive-by Teachers." *Carnegie Perspectives*. Palo Alto, CA: Carnegie Foundation for the Advancement of Learning, June.

What Works Clearinghouse. (2005). www.whatworks.ed.gov.