Assessment of Basic Math Skills
at the University of Wisconsin–Platteville

Barb Barnet
Dave Boyles
Kevin Haertzen

Department of Mathematics
University of Wisconsin–Platteville

December 5, 2005


**Introduction**

In the past few decades one of the more useful tools of educational reformers has been assessment, which has grown into a popular method for the exploration of the possibilities of teaching improvement. There is a long history in education of assessing the impact our educational practices have on students. In some sense, assessment is what we are doing every time we grade a student’s work or hand out grades at the end of a semester. Before a student walks across the stage at graduation, an assessor checks whether they have met the requirements for graduation from our school. But assessment is more than that.

*The higher education community as a whole seems to have accepted accountability as a positive concept. Based on my conversations, assessment is being seen more and more as a means to access larger issues of institutional purpose and to make progress towards these goals, rather than the narrower concept of assessment as a quality measurement at a fixed point in time.*

Accountability is important, however a danger is that discussions threaten to devolve into questions about how we can most easily meet our various requirements. And then, it is hard to argue that the results of all our hard work actually mean something. If we want our assessment to be meaningful, it must be grounded in a desire for improvement. Improvement might be suggested in the structure of the curriculum, in the methods of teaching that faculty are using; it even can be of the assessment itself. In any case, improvement is the point, and the ability to meet accountability standards is just a welcome by-product of the effort.

*The many facets of assessment can be summarized in the recognition that we are still imperfect teachers. In the great balancing act between what is desirable and what is practicable, there are always opportunities for us to do a better job within the parameters in which we work. Assessment is the continuing process of finding that better balance point.*

1. **Meaningful assessment**

We see that assessment is a two step process: 1) attempting to quantify the impact of an education on your students at your particular institution of learning and 2) attempting to use the data to inform discussions on steps a particular curricular unit should take to improve the next set of numbers. Our purpose here is to describe such a process with which information can be gathered and be available for feeding back to the relevant policy-making university or department committees.

Thus, we are assuming that the accountability requirements have been met elsewhere. Assessment here is meaningful insofar as it illuminates discussions of curriculum. We propose three criteria one uses to judge how well our assessment might succeed in this, pointing out that each criterion exists on a fuzzy continuum from best to worst and one can only reasonably expect that the methods one chooses are more good than bad.

First, there is a need for your assessment tasks to be **authentic**. Inauthentic assessment mostly consists of simple substitutes for more direct examination, from which we want to make inferences about student abilities to perform, without actually seeing them perform. Authentic assessment directly evaluates performance on intellectual tasks that we value. Metaphorically, you don’t ask students: *which tool is this?* — you ask them to build you something. Keep in mind that inauthentic assessment is not necessarily useless. For example, if you wanted to assess someone’s skill as a truck driver, you would ask them to drive a truck for you. However, it would also be good to have them sit down and take a written test of their knowledge of the rules of the road; somewhat inauthentic, but quite necessary in the big picture.

The second criterion is that we want to our assessment to be **deep**. We would rather not just pick up small pieces of what students have learned at the surface of a subject’s meaning. Surface learning is seen to be

---

1. [Moses]
2. [Bressoud], p.580
3. We also point out that other criteria exist that we are ignoring; e.g., we want to believe that our instrument to be **valid**, in that it is a reliable indicator of the assessment measure in question. This can be a difficult issue to face at test-writing time. *Ready! Fire! Aim!* is a good motto to live by. If you wait until your aim is perfect, you might never fire.
inadequate in that it aims to accumulate, without understanding, facts that are answers to simple questions. For example: what atoms make up a water molecule? An important question perhaps, but whether or not a student knows the answer is not particularly useful assessment information.

Deep learning, on the other hand, is grounded in thoughtful study and, with reflection, leads to mature understanding.

*Deep learners can self-correct, consider a variety of perspectives, understand the larger context surrounding a particular problem, and relate theory to practice.*

A question that goes deeper than the previous might be: *what are the implications of the fact that water is at its most dense at four degrees above freezing?* One might imagine the responses that such a complicated question might elicit.

You want students to understand the subject deeply enough that they might recognize it in their experience of everyday events.

*The new teaching/learning paradigm includes a move away from knowledge-based learning outcomes that focus on facts and principles to “deep learning” outcomes such as the ability to apply knowledge and critical thinking and other skills. Deep learning should be important to us all, because employers, legislators, other key constituents and the students themselves all want us to develop students’ abilities; it’s what makes a college education worth the investment. But even more important is that today’s world presents incredible challenges that won’t be met by people who have superficial educations. The task before us is to be able to assess the complexities of deep learning.*

The problem with assessment of deep learning is that a good tool is not easy to construct. And, more to the point, that mathematics is at a disadvantage on this count (at least until the day we start putting essay questions on math assessment tests).

A troubling issue that arises is what your reaction should be to students’ inability to perform on a simple task that you consider to be essential knowledge. Suppose that a large majority of your sample does not know that there are two atoms of hydrogen and one of oxygen in the proverbial water molecule. Is this a problem? Doesn’t everybody already know the answer in high school? If it is a problem, what would you propose to remedy the situation? Does that mean we must include the correct answer in all the science classes the university teaches? Now suppose the same questions for a more complicated molecule like nitroglycerin. There must be a line here somewhere, with nitroglycerin on one side of the line and water on the other. And, even if you don’t believe that knowledge is the key ingredient — that skills possessed by the learner are much more important — the proposition remains: skills on the water molecule side of the line are so important that we must not only teach them; just as importantly, we must assess them.

Third, we want our assessment to be **effective**. That is, when the tasks are complete, the scores compiled, and the statistics generated, we want to have useful information to feed to the various curricular bodies that is specific and reliable enough that it may be used to inform policy decision-making. This can be a daunting task. Even if we have satisfied our expectations under the first two criteria, there is no guarantee that we can even come close on this last criterion. Yet without the feedback step, we find ourselves merely going through the usual assessment motions, perhaps convincingly enough to satisfy accreditors, but unsatisfactorily nevertheless. In particular, where is the incentive to do it all again two or three years hence?

For assessment to have an effect, foremost in preliminary discussions must be the mission of the school in which one works. Its educational values should dictate not only what is assessed, but who is assessed and how we decide to do so. When mission and values are ignored, assessment becomes an exercise in studying what seems easiest to measure. We end up throwing out the baby and keeping the bath water, simply because we can measure how dirty bathwater is.

---

4 [Prados]

5 Also, one notices that such a question is not appropriate for a multiple-choice test like the one we will be writing.

6 [Angelo, Ewell, Lopez]
To guard against this mistake, first carefully articulate the standards that we are measuring the success of our efforts against. These standards should be authentic, deep, and in harmony with our mission; but most of all they must be assessable. In other words, they should not try to describe the well-educated student as much as they should be a list of actions that we would think an educated student should perform when faced with the kinds of choices that they would have to make after graduation.

The standards building step in the process is in many cases the easiest to get colleagues involved in. Discussions among educators regarding the fundamental meanings of the subject that they teach can inspire passionate examination of their most deeply held beliefs. It has been said more than once that the simple fact that such dialogue is made possible already validates the worth of assessment as a pedagogical tool. No such claim is made here; in fact, the danger is that the process can sink into the bog at this step, and we would like to show that assessment is not only interesting, it can be effective.

2. The basic math skills

At UW-Platteville, the teachers of the Department of Mathematics are faced with four types of students: beginning engineers, math majors, prospective K–12 teachers, and “captured” students looking to meet their mathematics competency requirement. The purpose of the assessment exercise that is the subject of this paper is to measure our success at meeting the needs of this last set of students. Are we teaching them to be numerate (i.e., quantitatively literate)?

Most university faculty that you talk to will agree that there is a definite need for some sort of mathematics competency requirement for graduation. We can assume that this is the case, noting only that the fact that it is true is not as obvious as one might think (and could be the subject of a separate paper).

To cope confidently with the demands of today's society, one must be able to grasp the implications of many mathematical concepts — for example, change, logic, and graphs — that permeate daily news and routine decisions — mathematical, scientific, and cultural — that provide a common fabric of communication indispensable for modern civilized society. Mathematical literacy is especially crucial because mathematics is the language of science and technology.

The idea that all college graduates should be expected to have acquired a certain familiarity with mathematics rests in part on the well-founded belief that such a familiarity is necessary for effective functioning in contemporary life.

But when it comes to standard-writing time, committees spend a lot of time trying to describe the ideal student, from which grows a tendency to then set the bar higher than most students can jump. For example, the MAA: … every college graduate should be able to apply simple mathematical methods to the solution of real-world problems. A quantitatively literate college graduate should be able to:

- interpret mathematical models such as formulas, graphs, tables, and schematics, and draw inferences from them;
- represent mathematical information symbolically, visually, numerically, and verbally;
- use arithmetical, algebraic, geometric, and statistical methods to solve problems;
- estimate and check answers to mathematical problems in order to determine reasonableness, identify alternatives, and select optimal results;
- recognize that mathematical and statistical methods have limits.

---

7 …the wide variety of students who, much to their surprise and disappointment, are suddenly forced to study a subject that they have been fleeing for years, even if fresh out of high school. [Stein]: p. 1023

8 [Paulos]

9 [MAA], Part 1: Why Quantitative Literacy?

10 [CUPM]

11 [MAA], Part 2: Quantitative Literacy Goals
In any case, the MAA sets the bar equally high for the colleges and universities as whole: [all] **colleges and universities should:**

- treat **quantitative literacy as a thoroughly legitimate and even necessary goal for baccalaureate graduates;**
- expect every college graduate to be able to apply simple mathematical methods to the solution of real world problems;
- devise and establish quantitative literacy programs each consisting of a foundation experience and continuation experiences, and mathematics departments should provide leadership in the development of such programs; and
- accept responsibility for overseeing quantitative literacy programs through regular assessment.\(^\text{12}\)

Whether or not you believe that there truly are “quantitative literacy programs” at your university, the last bulleted item describes the charge an assessment committee is given.

So we need to write our local assessment standards, but before we do, we need to answer a question. In 1978, a national survey asked: *What mathematics should every graduate of an American college or university know?*\(^\text{13}\) Without looking below, the reader should attempt their own list, to see how in tune they are with the committee’s findings. The following were marked by at least half the respondents.

- **basic arithmetic skills** (94.6%)
- **area and volume of simple shapes** (76.4%)
- **linear equations** (71.3%)
- **algebraic manipulations** (63.0%)
- **elementary statistics** (55.5%)
- **graphing of functions** (54.9%)
- **exponents** (54.3%)
- **plane geometry** (51.9%)\(^\text{14}\)

These skills are expected to have been “mastered” by all college graduates; however, there is no reason to believe that one must teach them all. For example, basic arithmetic skills scored extremely high (94.6%), yet not very many would propose that college courses should be teaching them, even though a teacher of college mathematics often finds oneself teaching to students skills they should have learned in high school.

A slightly different question might be asked: *What basic mathematics skills should every person, who is considered to be educated, have?* In other words, are there certain fundamental notions in our subject that are so important that society would be a better place if most adults understood how to work with them? The mathematics framework of the Nation’s Report Card on Mathematics includes five “broad strands” of mathematics content:

- **number sense, properties, and operations;**
- **measurement;**
- **geometry and spatial sense;**
- **data analysis, statistics, and probability; and**
- **algebra and functions.**\(^\text{15}\)

So, if we include measurement and probability/statistics under the general heading of *data*, we see that there are four basic emphases: **numbers, space, data, and algebra.**

---

\(^{12}\) *Ibid.*, Preface

\(^{13}\) \[CUPM\]

\(^{14}\) Others with less than 50% include: **probability, problem solving, quadratic equations, radicals, business applications, logic, computers, and systems of equations.**

\(^{15}\) \[NAEP 2005\]
• **Number sense**: operations, properties, relations, patterns, counting, …
• **Spatial sense**: areas/volumes, shapes, transformations, invariance, graphs, …
• **An ability to work with data**: measurement, estimation/errors, probability/statistics, technology, …
• **An ability to work with mathematical certainty**: algebra, functions, logic, proof,…

The first three bullets we shall call the basic math skills. While the fourth is important, it is essentially the list of skills a student needs to develop in order to move into learning mathematics at the calculus level and above. We compare with the following: from the 2005 UW-Platteville General Education Plan.

*Students should have a basic competency in computational skills and quantitative perception. A course meeting competency requirements in mathematics is designed to enable students to:*
  • develop problem solving skills using the methods of mathematics;
  • use the recognition of patterns to solve problems;
  • work with fundamental notions of number and space;
  • distinguish between valid and invalid reasoning; and
  • remain alert to the plausibility of solutions.*16*

These are really not too bad; however, they are not assessment standards. For that reason, it doesn’t seem all that clear what anybody is supposed to do about them, even if they are in full agreement with the cause. Note that the last two bullets are as much critical thinking skills as they are math skills, but they still belong on our list. The next step would be to turn the above educational standards into a short list of assessable student learning outcomes, in time for the next round of testing.

### 3. Creating the local instrument

*The bottom line is that the purpose of any mathematics assessment must be to improve students’ learning. When an assessment measure is well aligned with — and integrated into — the system of mathematics and teaching and learning, preparing students to perform well should involve little more than teaching the mathematics program well.*17

Some of the deeper issues in our basic skills program were being discussed in the UWP math department in the fall of 1994. At that time we needed to respond to what was considered to be a flawed UUCC assessment of our basic skills courses. Subsequent discussions indicated that that there had been a desire for some time, in various governing bodies at UWP, that local assessment strategies and instruments be developed. In the years before the last accreditation visit of North Central — now the Higher Learning Commission — there was pressure to create a home-grown instrument for the assessment of student math skills. So the question became: what should we use to measure the skills of our students?

Testing is usually the first option discussed when actual practices come into the conversation, especially in a discipline as formal as mathematics. The reasons for this are good. Tests are inexpensive compared to other choices that might be expected to directly improve student learning. Tests can be, often quite thoughtlessly, externally mandated. Results can be certainly reported to the press (at least the promising ones). But there are side effects.

*The instrument can be capable of a diagnosis of educational excellence, but most school systems have no experience in using data in a real-time way to inform instruction. The tests are not used as a snapshot; they are a goal.*18

Here the author has given us an insightful definition of successful assessment: *using data in a real-time way to inform instruction.* This describes well the ultimate goal that we hope the many efforts described in this report

---


17 [Seeley]

18 [Hodas]
are moving us toward. And so a testing strategy must be found. As long as one is simply concerned with creating a multiple-choice math skills assessment test, the task is fairly simple.

Furthermore, one can argue that a multiple-choice test, whose questions require the test-taker to choose from a list containing the correct answer and (say) four plausible distracters, is not necessarily inauthentic. As in life, where one is often faced with questions that have more than one possible answer, the real question is: which one is correct? Many of the problems on your test will be solved by students simply calculating the answer and then finding it in the list of choices. However, again as in life, many problem situations present various possible solutions and the problem is largely one of choosing well. Many mistakes are not a result of failure to recognize the correct solution as possible; instead one fails to recognize the correct solution as best.

Thus, authenticity is — at least partially — achievable. And, with some effort, one can compose math problems that assess some of the deeper learning that we are interested in measuring. On this point, it is important to have, on your assessment committee, colleagues who understand both the assessment standards and the reason for the content of the courses in the General Education Plan.

Nevertheless, a subtle issue arises sooner or later, and stops you in your tracks. To explain this, we need to first distinguish between two types of assessment:

- **formative**: assessing progress during the learning process to give teacher and learner feedback, and
- **summative**: assessing terminal performance to see if the learner can meet a standard.

A suggestion box is a simple example of formative assessment, while the SAT exam is summative. Taking a truck driver’s exam is a summative assessment of a beginning driver who has completed a training course, while an attempt at a formative version might be a sign on the back of the truck that reads: *How’s my driving? Call 1-800-xxx-xxxx.* (Not that anyone ever calls that number — a common flaw in perfunctory assessment.)

A summative math skills assessment instrument is easy to conceive of and not really all that difficult to execute. So, it is no surprise that the product of all the preliminary philosophizing is exactly that. However, in order to make this assessment effective, you are looking to use your results formatively. There is an inertia that arises after completion of a few rounds of testing: nothing seems to want to move. The subtle issue that your assessment committee faces is the timeless square peg into round hole syndrome; trying to use a summative experience in a formative way. You find yourself reporting to some governing body what percentage of students got question \( n \) right, and really don’t have an answer when someone asks: so what?

Recognizing the pitfall that undoubtedly lies ahead, one moves forward anyway (ready, fire, aim!). We report here that: funded by assessment grants, we have created the home-grown math skills test recommended by the Assessment Oversight Committee all those years ago. It is important to understand that the twenty-question test itself is only a small (though vital) part of the project. The organization, delivery, and reflective analysis involved requires the focused attention of a group of colleagues over an extended time period. In the end, the assessment we have created is more authentic, deeper, and most importantly, more relevant to the students taking their UWP math competency course than the ACT-CAAP can ever be.

The first and third exams were given in 1998 and 2001, during the Assessment Oversight Committee’s initial administration of the CAAP test. On the other hand, the second run-through in 1999 was with students who were completing their required math general education course. Although this turned out to be harder to create the sample of test-takers, this still seemed in the end to be a better method. In particular, there is the possibility of formative feedback. It is instructive to note how the same twenty question test can change from completely summative to somewhat formative simply by changing the administration of the instrument.

---

19 A distracter is an answer that is incorrect but can be found by “almost” solving the problem correctly. One finds that the best test questions are the ones with distracters that can be explained by a fundamental mathematical misconception. For from them, we are able to infer what the students may be thinking as they work the problems.

20 And, there are excellent standardized exams like the ACT-CAAP, where all a school needs is money and volunteers.

21 In 1997, from the deep pockets of the Provost, and in 1999 and 2001, by the Assessment Activity Fund.
Therefore, when time for the fourth round of exams approached in the spring of 2004, the department assessment committee decided to use the method of the 1999 test.

4. The 2004 exam

In the spring of 2004, 238 students enrolled in the courses of Finite Mathematics, Mathematics of Finance, Elementary Statistics, Trigonometry and Analytic Geometry, and Calculus II took the Math Skills Test. Another 91 students in Mathematics for Educators II and Pre-Calculus took the exam in the fall of 2004. The table below displays the breakdown of the number of students who took the exam in each course, as well as the total number of students enrolled in that course during the semester the exam was administered.

Within each course, anywhere from 19.7% (Elementary Statistics) to 81.4% (Trigonometry) of the total students enrolled completed the Math Skills Test. Although only around 20% of the Elementary Statistics and Calculus II students took the MST, these courses have very large enrollments and the number of students who took the test is comparable to those of the other courses. Even though the exam was given during two different semesters, none of the students completed the exam more than one time. Also, a close examination of the answers revealed none of the exam results needed to be discarded due to lack of answers, patterned answers, or other problems. This resulted in a total of 329 exam results.

<table>
<thead>
<tr>
<th>Course</th>
<th>Test Date</th>
<th># of students who took math skills test</th>
<th># of students enrolled</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>163: Finite Mathematics</td>
<td>4/30/04</td>
<td>14</td>
<td>22</td>
<td>63.6</td>
</tr>
<tr>
<td>173: Math of Finance</td>
<td>4/27/04</td>
<td>54</td>
<td>86</td>
<td>62.8</td>
</tr>
<tr>
<td>183: Elementary Statistics</td>
<td>4/22/04</td>
<td>50</td>
<td>254</td>
<td>19.7</td>
</tr>
<tr>
<td>203: Math for Educators II</td>
<td>12/02/04</td>
<td>30</td>
<td>38</td>
<td>78.9</td>
</tr>
<tr>
<td>245: Pre-Calculus</td>
<td>12/13/04</td>
<td>61</td>
<td>149</td>
<td>40.9</td>
</tr>
<tr>
<td>253: Trig. / Analytic Geometry</td>
<td>4/26/04</td>
<td>70</td>
<td>86</td>
<td>81.4</td>
</tr>
<tr>
<td>274: Calculus II</td>
<td>4/30/04</td>
<td>50</td>
<td>235</td>
<td>21.3</td>
</tr>
</tbody>
</table>

One of the primary purposes of administering the 2004 MST was to investigate the utility of the questions that were used. One of our concerns was whether any of the questions had results that indicated that a substantial number of students guessed at the question and we also wanted to look closely at any questions where more students answered the questions wrong rather than correctly. Two questions that stood out were 2 and 20 (displayed in the graphs below, with correct answers marked by an asterisk; to read the questions, see the exam attached to the end of this report).

The suggestion has been made that problem 20 is inappropriate because of its difficulty, which may or may not be true. However, question 2 is appropriate, is worded quite clearly, and is basically a very simple question. So what went wrong? If nothing else, we have identified a problem with our general education standards: we say that we expect students to work with numbers (and how many elements a set has is a fundamental numerical problem), yet something is wrong. Whether this problem should be fixed is a question that might be examined — examination such as this being the ultimate purpose of any assessment exercise.
There were questions, for which the proportions of right vs. wrong answers depended on the particular gen. ed. course the students happened to be enrolled in. To illustrate this, we separate the courses; e.g.,

3. If a circle of radius $R$ has the same area as two circles of radius 6, what is $R$?

A. $\sqrt{2}$  
B. $\sqrt{6}$  
C. $\sqrt{12}$  
D. 12  
E. $\sqrt{72}$

This question has appeared on our test in some form each time we have given it. Its first incarnation was a question about two oil pipelines and the Environmental Protection Agency, which was essentially the above question. It was decided this time around to ask the fundamental mathematical question, without clouding the issue with a context. So, the question we ask of the test-taker is: of the five circles (drawn below approximately to scale), which has the total area of the two circles above?

In the histogram, the answers are all over the place: for example, almost 40% of the 203 (Math for Elementary Teachers 2) students think the answer is B, which is a very poor (way too small) guess, while over 70% of the 274 (Calculus 2) students have it right. The simple truth certainly is that very few, if any, of the non-274 students had rough sketches like the above five circles in front of them — on paper, or even just in their imagination. To an experienced math teacher, this is not surprising.

It should be clear that answer D, $R = 12$, is much too large; for example, you can fit two circles of radius 6 into a circle of radius 12, with a lot of area uncovered.

It seems that the answer must be somewhere between 6 and 12, for which E is the only candidate. Yet more students picked D than the correct answer, E, in every course except 245 (Pre-Calculus) and 274. Note that 253 (Trigonometry and Analytic Geometry) is not one of the exceptions, a somewhat surprising result given the content of that particular course.
Other questions with a disproportionate number of wrong answers include 6, 14, and 18. At the other end of the spectrum were relatively easy questions, of which the two easiest were 1 and 17 (below).

Question 1 is a very simple problem, included for no other reason than to get the test takers off to a good start. Near the end of the test, the reason for including 17 was to see if we could imagine how many students worked all the way to the end of the exam. Almost 90% got either the correct answer D, or missed the change of units in the question and got A, which is an honest mistake. We conclude that most of the test takers gave the exam a good effort.

One notes that even an easy question can be hard for some students. For example, questions 7 and 9 had similar patterns of correct answers, with the 163 students doing poorly, the 274 students quite well, and every other class in between. The exception was 203 students, who picked the correct answer more than 70% of the time for question 7, while scoring less than 30% correct on question 9.

The interested reader can probably put forward a cogent explanation for the difference, based on the differing nature of the particular questions.

The point to be stressed here is that no information like this was available for the years we gave the ACT-CAAP. This was for at least two reasons: 1) we had no way of seeing how the students were doing, question by question — even what the questions were, and 2) there was no useful way to break the student sample down into subcategories like we have done here. There was an attempt on the part of the Assessment Oversight Committee to create regressions using variables like ACT exam scores or math placement levels. The problem with trying to do this ourselves is that the relevant student data is not easy to get your hands on (and maybe not even legal to do so, at least in some cases).

5. Student performance in second-semester calculus

Math 274 is the second semester calculus course at UWP. In the spring of 2004, Kevin Haertzen was an instructor for two sections of Calculus II. He collected and graded homework on a weekly basis, which counted for 25% of the students’ final course grade. In lieu of any homework assignment, students were allowed to replace their lowest homework score with a 100 (out of 100) if they took the math skills test. Participation was completely voluntary, however most (49 of 54) students took the test, which was administered on the same day to each of the classes.
The majority of each 274 class consisted of traditional engineering students, who are required to pass three semesters of calculus and one semester of differential equations. Although the MST is not designed to test competency in calculus, we felt it would give us a benchmark with which to judge the lower level students who were taking the test. The test was administered to 49 (of 54) of Dr. Haertzen's Calculus II students in the latter portion of the spring 2004 semester.

We took the opportunity to compare their results on the assessment test with their final grade. Nine of these students earned a course grade of D or F. Their results on the skills test ranged from 11 to 17 correct. The remaining 40 students passed the course and their scores ranged from 6 to 17. Students who failed to pass Calculus II did not necessarily score lower than those who passed the course. An inference one draws is that something other than lack of basic skills preparation explains their failure to pass the course.

It was expected that, as a group, that these students would score higher on most questions than any of the other classes did. Success in a second-semester calculus course requires the utilization of material from the previous calculus semester, e.g., limits and derivatives. Students at this level exhibit strong analytical abilities. Learning about some of the standard methods of integration and working with series and sequences helps to further develop their problem-solving skills. Not all students will succeed in the course, but they are all in an environment of high expectations and amongst students who are aggressive in their mastery of the material. And, in the end, that these students did score higher on most questions.

Even though the problems in the MST are not selected from the usual calculus topics, the more mature the mathematical understanding of a student is, the more likely it is that they will be able to discern what the question means and understand what the answer might be. In particular, one expects a 274 student to avoid the various innumerate distracters that appear in certain problems.22

Because the students in this sub-sample are more mathematically aware than the average gen. ed. student, one expects that 274 would have the highest percentage of correct scores of all classes — and expects 245 to finish second most of the time. This is the result for all but a few questions; for example, question 3 above (where the only courses for which the correct answer won the vote are 245 and 274).

Those few questions where 274 was not either first or tied with 245 for the most correct answers, are interesting, if only to question why. Question 5 stood out in this regard (above left). Even more striking than the fact that only 40% of the 274 students had it right, is that 60% of the 203 students had the correct answer.

A second such question was 19 (above right): most striking is that the 183 (Elementary Statistics) students tied with 274 for second place. One notes that the 274 percentage dropped for the last five questions, which is explained partially by the increasing difficulty of the questions, and partially by the students starting to lose interest. A challenge of assessment is to get students to take it seriously, for the entire test.

A student learns to think about problem solving, the way a mathematician would, by solving lots of problems. Doing problems that were nothing like the next test question helps the mature student to see possibilities for solving it. Metaphorically, you teach a student to be a music instructor, in part, by having them learn to play something. The particular instrument chosen is not that important — what is important for them as a future teacher is that they can learn to play it. For deeply learned knowledge is transferable.

---

22 After the 2001 exam, a list was made of the grossly incorrect answers on the test. On breaking down the test results by college, it was found that EMS students chose one of the various innumeracies as their answer almost none of the time. So it turns out that, although engineering, math, and science students did make mistakes, they were not gross mistakes.
6. Conclusion

Doing assessment is not giving a test, watching a student performance, or handing out a survey to employers our alumni work for, although each is a pretty good idea. Part of assessment is composing a plan, but having a plan is not assessment. It is certainly not reading a number on a sheet of paper and then putting the sheet into a desk drawer. So what is assessment? And, more importantly, are we doing it? 23

An analysis of the issues that an institution faces trying to create its own authentic assessment process, listed the following five criteria for deciding whether an assessment plan can be effective.

- Analysis of assessment results
- Establishment of a clearinghouse for assessment information
- Communication of assessment results
- Establishment of a formal policy re: use of assessment results
- Documentation of the use of assessment results 24

Doing assessment well can be thought of as keeping an eye on all five bullets, especially the last. At the close of their input into the above case study, each institution was asked what the most important advice they could give readers. The majority of responses were essentially one of following two statements.

- Have the support of senior administrators
- Fit the implementation process to the way the institution currently functions 25

We emphasize the essential relevance of the second bulleted statement, and note the fact that as the MST was being developed, the advice was understood.

A colleague recently described about grading exams in the order that they were handed in, making sure not to shuffle the order at any point. This was to see if there might be a correlation between how much time a student put into the test and the grade they got. Assessment is coming up with ideas like that, in a structured, systematic way, and then putting the knowledge so gained to good use in your teaching. In a fundamental sense, assessment is simply that state of mind in which various truths, which in a more perfect world would go without saying, are repeated again and again.

While it is interesting that it can be reported to external accreditors that the math department is now assessing the basic skills, it would be far more effective to be able to affirm, e.g., the “establishment of a formal policy re: use of assessment results”. The exam here is in the end only summative, but it can be used formatively if the way has been prepared first. The next step would be for the math department to begin to articulate our assessment standards, preferably in the language of specific outcomes. 26

7. Recommendations

- The math skills test should be repeated, with the two or three least effective questions replaced.
- The math department should provide the assessment committee with a list of a few (3–6) math skills learning outcomes. These are to be used as goals for the test in bullet one. Possible benchmarks will be found in a more comprehensive list of statistics re: the 2004 exam.
- A report on this paper should be presented at a conference. With HLC re-accreditation on the horizon, there should be funds for this.
- Someone should submit this for publication. The assessment community needs to hear about this.

23 See [Steen ARQ] for an excellent discussion of the issues involved.
24 [Nichols], p.50
25 Ibid., p.73.
26 For example, at least 40% of the test-takers will get at least 9 answers correct — not a very useful learning outcome, but the point of the example is clear.
8. References

[Angelo, Ewell, Lopez]: Thomas Angelo, Peter Ewell, Cecilia Lopez, *Assessment at the Millennium, Now What?*, plenary session at the 1999 AAHE Assessment Conference


