Dan Kennedy

Assessing True Academic Success: The Next Frontier of Reform

Suppose that your doctor's idea of a successful medical practice is to teach his patients how to pass every diagnostic test that he gives to them. How will you ever know whether you are healthy? This question is a facetious one for doctors but a serious question for teachers.

As a teacher, I must admit that I never really thought too much about the impact that my tests were having on the course that I was teaching. I did realize that tests were important, because a student had to *do* mathematics to really learn mathematics. I also realized that no matter how much my students seemed to appreciate my classroom exposition, only in testing them would I discover whether they truly understood. Tests, in other words, were the validation of the contract between teacher and student; they were the proof that learning had taken place. That approach was fine with me, and even though my students did not always appreciate my tests, the general premise seemed to be fine with them.

And so it went for roughly twenty years of my teaching career. Then I became involved in writing problems for the Advanced Placement (AP) calculus examinations, and before long I was frequently confronted by my colleagues about various issues of assessment.

One of the most persistent complaints about the AP examinations from college professors in the early days of calculus reform was that they were too predictable. Enough variation occurred from year to year to enable our committee to deny that we worked from a precise template, but if we deviated too far from what high school teachers thought that template was, they would complain. We could leave off a particular type of problem for one year with few repercussions, but if we left it off for two consecutive years, AP workshops all over the country would have to explain whether this omission meant a change of direction for the program. Teachers believed that they had a right to know what would be tested, and why not? It was their job, after all, to prepare their students for the test. Moreover, since the test would be the only measure of their students' success that would matter in the end, the extent to which calculus was both taught and learned in their classrooms would be, like it or not, defined by that one experience on that one morning in May. Faced with that stark reality, teachers were reduced to uttering in dead seriousness the six words that they hated most when emanating from the lips of their students: "Will this be on the test?"

At about the same time that I was being challenged about the predictability of the AP examinations, I began to worry about the extent to which I, rather than the students, was doing all the interesting mathematics in my own classroom. I had been railing for years against the way that students seemed to lose the ability to think as they progressed further and further through high school, but I had never suspected the extent to which we had been drumming it out of them by making them play our educational game. The rules of that game are simple: we, the teachers, show them what to do and how to do it; we let them practice it for a while; and then we give them a test to see how closely they can match what we did. What we contribute to this game is called "teaching," what they contribute is called "learning," and the game is won or lost for both of us on test day. Ironically, thinking is not only absent from this process but in a curious way actually counterproductive to the goals of the game.

Dan Kennedy, dkennedy@chattanooga.net, teaches at Baylor School, Chattanooga, TN 37401. He is interested in all aspects of educational reform, especially as it involves Advanced Placement mathematics.

Photo credit Comstock, Inc.

Photo credit Comstock, Inc.

Think about it. Thinking takes time. Thinking comes into play precisely when we cannot do something "without thinking." We can do something without thinking if we know how to do it really well. If students can do something really well, then they have been very well prepared. Therefore, if both we and our students have done our jobs perfectly, they will proceed through the test without thinking.

If we want students to think on a test, then we will have to give them a question for which they have not been fully prepared. If they succeed, fine; in the more likely event that they do not, then they will rightfully complain about not being fully prepared. We and the student will have both failed to uphold our respective ends of the contract that the test was designed to validate, because thinking will have gotten in the way of the game!

Thinking is a creative act. Thinking and learning ought to go hand in hand, but many of the things that we learn to do are not dependent on creativity, so thinking is not really involved. We learn how to walk, how to tie our shoes, and how to ride a bicycle. On a slightly higher intellectual plane, we learn how to read, how to write, and how to do arithmetic—a noncreative act that I do not count as mathematics even if many other people do. These skills are things that we learn how to *do*. We simply learn other things, such as history, appreciation of Shakespeare, or quantum mechanics. The extent to which these things are learned seems to be proportional to the extent to which we think about them, but they are not things that we *do*. Indeed, once we get past the primitive skills of reading and writing, only one classical intellectual pursuit in our academic lexicon is relegated to the status of learning how to walk, ride a bicycle, or tie our shoes. We learn *about* everything else; we learn how to *do* mathematics.

When educated adults meet us at parties and chuckle that they "never could do mathematics," they are not talking about arithmetic, nor are they talking about thinking and learning. Educated adults do not forgive themselves readily for not thinking or for not learning. They are talking about a noncreative act that some people do, like juggling, and that they, by fate or by preference, simply do not. Ironically, these same people might have no problem balancing checking accounts, comparing stock portfolios, cutting recipes in half, estimating their gas mileage, computing restaurant tips, or figuring out how long to cook a twelve-pound turkey. We might call that doing mathematics, but they do not, especially because they can do it. What they could not do was defined for them long ago, in those tests at the end of the game, and test performance was what our educational game had defined as doing mathematics. Some teacher had shown them repeatedly how to factor, and when the time had come for them to show it back, they had failed. And that forever was that.

We might think that science would be in the same fix as mathematics as far as these perceptions are concerned, but physics, chemistry, and biology are still subjects primarily to be learned *about*, at least in high school. When we do an experiment in chemistry, we discover something about how the world works; we are playing a creative role in the learning process. Besides, even in high school, science is a living subject. Nobody talked about big bangs, plate tectonics, recombinant DNA research, AIDS, black holes, or quarks when I was in school, because my teachers and my textbooks, through no fault of their own, did not even know what they were. Like most people in my generation, I have learned about these topics since high school. My science teachers, I suppose, would be proud, except that they would point out that I am hardly a special case. However, anyone who admits to having learned something about mathematics since high school is a special case indeed. In fact, not many people would admit to having learned anything about mathematics in high school either; they will say only that they were good or bad at *doing* it.

So to get back to my own thinking, I began to realize a few years ago that the mathematics in my own classroom was not a living subject; that it was not a creative, thinking act for my students; and that they did not view it as something to be learned but rather as something to be done. No matter what else I did, the bottom line was that I should show them how to do it so that they could do it back for me when the time came. In other words, we were both playing our roles in the game to perfection.

That was when I decided to try not playing the game.

I began starting each class with a problem rather than with a monologue at the chalkboard. Then I would walk around to see how students solved it in collaboration with one another. Sometimes I had to do a little coaching, but they eventually discovered the mathematics themselves. Then we talked about it. If I needed to, I would give it a name or a historical context. If some subtlety would affect the solution, I would give them another problem and watch them deal with the subtlety. Then they would explain it. Quite often, they came up with alternative solutions, so we talked about those. They were still *doing* the mathematics, but now they were learning about it and thinking about it at the same time. Best of all, the students could finally appreciate the need for creativity in really doing mathematics.

Homework became an extension of the classroom experience—a continuation rather than the beginning, of their own performance. When my students came upon a problem that was unfamiliar to them, their first instinct was to learn how to solve it rather than to blame me for not playing my part in

Anyone who admits to learning something about mathematics since high school is a special case indeed

the game successfully. I found that they even read the book without being told to! (Has a parent ever told you that "last year's teacher was so good that my child never had to open the book"?)

Only after I had thoroughly quit the game did I begin looking critically at my old tests and wondering what good I could ever have imagined coming from them. They had obviously been designed to push the computational envelope, to see what the students could do with the skills that I had taught them, but they were consequently lacking in creativity and were no longer exemplary of what I was trying to get students to do in my postgame classroom. The tests would obviously have to change, but how far could I go in changing the rules of the game? How could I demand creativity on a test for which I had fully prepared my students, and how could I not fully prepare them when our mutual success depended on their performance? This dilemma seemed almost insurmountable until it occurred to me that it was a matter not of assessment but rather of grades.

I was determined not to let grades stand between my students and their learning

Since the goals of assessment were ideally independent of grades, I supposed that in an ideal world, grades would be irrelevant. However, I did not teach in an ideal world; I taught in a competitive prep school from which students expected to graduate and attend their colleges of choice. Students, parents, administrators, and colleagues not only saw grades as relevant but believed in their hearts that they were the most significant output of the so-called assessment process. In the dark and hidden depths of my own reform-minded bleeding heart, so did I. But my experience with mathematics had taught me that numbers should be our servants in applications, not our masters, so I was determined not to let grades stand between my students and their learning if I could help it. I also realized that I could obviously help it, since I controlled the grading algorithm.

Grades at any school lie along some numerical or alphabetical continuum that is, by itself, fairly meaningless. Attaching meanings to certain cutpoints along the continuum is therefore necessary so that the absolute worth of a grade can be interpreted by an interested observer, such as a parent or a college-admissions committee. The grading continuum in use at my school happens to be a percent scale, from 0 to 100. The failure cutpoint is at 65, effectively eliminating about half of that scale for any practical purposes. Does this scale mean that a student should be successful on 65 percent of my test to pass? Let us think for a minute.

No major-league baseball player has ever come close to batting .650 for a season. The best basketball players in the country shoot less than 65 percent from the field. A salesman who makes a sale two out of three times is a miracle worker. A student in AP calculus who solves 65 percent of the problems correctly will most likely earn a 5, the highest possible score. These people are experts in their fields. How can we justify demanding 65 percent from mere learners to demonstrate minimal competence? And if a beginner who is minimally competent can handle 65 percent of my test, what kind of competence could that test possibly be measuring?

My theory is that we ought to present students with challenging, relevant, useful, and varied assessments all the time and then scale the grades to conform to our expectations. We can accomplish this feat; we are mathematicians. I will share with my own method for scaling, although it might not be the best method for everyone. The only thing that I am advocating for everyone is that we all be freed from the tyranny of numbers insofar as they limit our freedom when it comes to assessing true learning.

I tell each of my classes on the first day of school their class average: 82 for a regular section of a required course, 85 for an elective, and 90 for an advanced section. These numbers are based on schoolwide empirical data. Whether I like them or not is as irrelevant as my opinion about the price of a first-class stamp and carries as much weight. I tell them that from that point on, they can raise the class average by exceeding my expectations and can lower it by disappointing me, but that class average will determine the scaling of my tests and quizzes. The better they are, the more they can expect me to challenge them—and the better will be their chances of showing me how high their class average should be.

So let us say that I give a challenging test to an advanced class whose average stands at 91. The students handle the material that I expect them to handle, and several of them surprise me on the hard problems. Grading on an absolute scale, I find that the test average is 75. I look back on the homework effort for the past few weeks, the class participation, and so on, and I decide to raise the class average to 92. I have an ordered pair (75, 92) for scaling raw grades to real grades. Suppose that my top student has managed a raw score of 93, a lovely paper, which I decide to scale to 99. I have a second ordered pair (93, 99). Those two points determine a linear equation that enables me to scale anyone's grade in a fair and objective manner. Mathematically, the effect of this scaling is to adjust the mean, a primary goal, and to reduce the standard deviation, a secondary effect that helps me keep my entire class engaged.

For example, let us suppose that this test really catches one student dismally unprepared, for any number of academic or other reasons. Say that the student gets a raw score of 20. My scale brings that score up to a 71, where it is still an outlier in terms of a much smaller standard deviation but where the student can still believe that a comeback is possible. The class average is very significant here: if the class average is 82 rather than 92, that raw score of

20 scales to a real score of only 30. This statistic has profound implications for the behavior of weak students in class!

Technology makes this linear scaling particularly easy to do. **Figure 1** shows a program that will work on the TI-82 and TI-83 graphing calculators.

The effect of scaling is dramatic. Freed from the shackles of unreasonable numbers, I can now challenge my students to do just about anything, then see how far they can go. They, in turn, have been freed from the burden of getting a certain percent right, so they can concentrate on doing as much as they can as well as they can. Moreover, they realize that the better they do as a class, the better the benefits of the scaling. In fact, my students and I see grades as measuring two things, both of them defensible on a loftier scale—namely, the quality of the class as a group and their relative standing within that group. If grades are to play a role in defining success at all, that approach seems to me to be a pretty good start.

We are still left with the matter of what we ought to be grading, or in a more general sense, assessing. A typical mathematics teacher probably currently assesses five qualities: knowledge, cleverness, diligence, context, and luck. I probably do not need to explain the first three virtues, except to note that knowledge, cleverness, and diligence are indeed quite different and that to be truly successful in most academic pursuits, a student should possess a generous helping of each. We ought to be assessing these qualities because as a society, we value them. The word *context* refers to miscellaneous other judgments that we make about a student's work on the basis of qualities that we value, such as neatness, punctuality, clarity of expression, elegance, and creativity.

Luck might appear to be a strange addition to this short list, but teachers assess luck all the time, whether they intend to or not. We intend to assess luck when we threaten students with pop quizzes or spot checks on homework. We indirectly assess luck when we give everyone in the class the same test on the same day—something that we all do. Luck is involved in being ready for a test, whether it involves emotions, health, the pages that the student studies, or what the student has for breakfast. We might not like to admit that we value luck, but it is hard to avoid testing it. The least that we can do is to try to minimize the effect of bad luck on our efforts to assess our students.

If we do value knowledge, cleverness, diligence, and various aspects of context, then we need to tell our students so, and we need to find as many ways as possible to let them show us what they have got. We should make a place in our assessments to test memorization because it is a part of knowledge, just as we should make a place for the problem that is deliberately intended to trick the knowledgeable student, as that is one way to assess cleverness. The diligent student should have a place to shine even if it might bore the clever one, and the clever student should have a place to shine even if it might dismay the diligent one. We need never apologize for holding students to such standards as correct spelling, typed essays, complete sentences, oral communication, or collaboration, because if we do not value context, then neither will our students.

```
ClrHome:FnOff
PlotsOff :ClrTable:ExprOff
6→Xmin:100→ Xmax
0→min:124→ Ymax
0→Xscl:0→ Yscl
Input "RAW SCORE: ",A
Input "CURVED TO: ",B
Input "RAW SCORE: ",C
Input "CURVED TO: ",D
(B-D)/(A-C)→ M
"round(MX+B-AM,0)"→ Y1
IndpntAsk
DispGraph
Text(1,1,"TRACE OR USE TABLE")
Text(7,1,"TO ENTER RAW")
Text(13,1,"SCORES.")
```

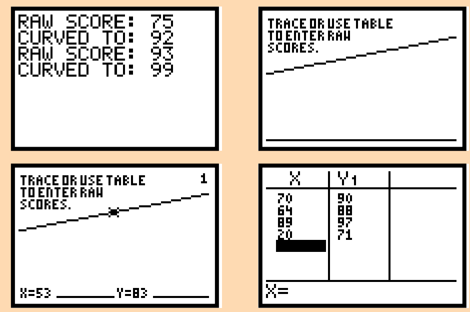

Fig. 1
A scaling program for the TI-83 or TI-82

For many years, my own assessments were heavily skewed toward testing knowledge. That game was the one to which I referred at the beginning of this article: the validation of the learning process, the definition of success. My major tests still focus on knowledge, but now I am careful to vary the context with algebraic problems, visual problems, straightforward and simple problems, problems that require writing, and problems that call for creative solutions. I assess cleverness primarily on weekly quizzes, which I tell students up front will not necessarily be "fair." I require collaboration in class and encourage it outside of class because I value it; it is how people learn all their lives. I give occasional collaborative quizzes and feel fine about assigning them grades. I can challenge the students even more on a collaborative quiz, and they do better work. Who loses in that deal?

We should minimize the effect of bad luck

I also have begun looking at portfolios. One of the most ironic aspects of traditional testing is that students cannot demonstrate what they have learned unless they are given the appropriate stimulus in a testing situation. This limitation not only introduces the element of luck, it also reduces the person who knows the most about what the student has learned—namely, the student—to a passive participant in the assessment of that learning. I require my students to keep a portfolio of eight items a year, and I specify only that each item should tell me something about the student's learning that I do not already know. A perfect quiz makes a bad portfolio item, since I have already

This year was my first year to be a peer tutor, and I enjoyed helping the girls in the dorm a lot. Last night, though, I finally saw the importance of my peer tutoring. My roommate came in at 10:00 extremely upset over her Precalculus test that was the next day. I calmed her down and told her that I would help her if I could. Carrie, who had been in the play, had gotten behind in her work, so she didn't understand what they were doing. She showed me the problem. I knew the answer, but I wasn't sure how to explain it to her in a way that was not confusing. I thought about it for a while, and I ended up trying several approaches (with Clara's help) that I had learned in Calculus, until I finally got through to her. Then I made her work a few problems for me, and she did them perfectly. She understood! I was so happy to be able to help her that I had forgotten I was supposed to be studying for my own Calculus test. She was so happy she understood that she began to cry. She really began to cry. It's great to be able to use the things you have learned to help other people learn too.

Fig. 2
Sample portfolio essay

given it a perfect grade. A bad quiz might make a great portfolio item if it alerts the student to some nugget of unlearned or mislearned knowledge that can then be learned correctly and reflected on in an introspective essay. I have learned fascinating aspects of the thought processes of my students by reading their portfolios. For example, **figure 2** is a brief essay that one girl, a boarding senior in my BC calculus class, wrote about helping her roommate prepare for a precalculus test.

Students have cried in my classroom before, but never from the sheer joy of learning. We often have no idea what kinds of learning occur outside our classrooms, nor can we fully appreciate the quality of that learning or the impact that it has on our students. At least with a portfolio, we have a chance to tap into that experience and include that evidence of learning in our overall assessment package. A footnote to that story is that Carrie scored 93 on that test the next day—a personal best and a full nine points above the class average. I believe that my calculus student enjoyed a richer learning experience helping Carrie with her precalculus problems than she would have had if she had studied for her own calculus test. By my feedback on her portfolio, I was able to affirm that learning experience, and I had no qualms about rewarding it with a good grade.

Nobody learns less mathematics than the student who stops taking it

With all these different kinds of assessments, is it possible to fail a course? Sure it is. But notice that a failing student must be bad at many things—bad with versatility, so to speak—and then must resist improvement in a variety of ways. As rare as it is to find a student blessed with abundances of knowledge, cleverness, diligence, and contextual excellence, I believe that it is even rarer to find a student who is lacking in them all and who can remain lacking for a year in spite of our best efforts.

A student who fights against failure is already a success by one measure; so why can we not find ways for that student to succeed on our tests, or more important, to learn mathematics? A student who surrenders to failure is failing far more than a course in mathematics; so how can we be accomplices to that kind of suicide? Nobody learns less mathematics than the student who stops taking it. We who are entrusted with teaching mathematics must consequently find ways to keep our students learning, and I am not convinced that failure is an effective strategy for any student in the long run. Let us face it: students who confront failure in classroom after classroom are the ones for whom failure becomes quite literally our long-run strategy; the problem is that such students never run with us for very long.

I must admit that I become a little impatient with people who say that we need to keep standards high by weeding out the students who cannot do mathematics and that to deny that premise amounts to sacrificing the mathematics in favor of student self-esteem. This matter is not one of self-esteem; it is about teaching mathematics. So, yes, I do try to keep my students around from semester to semester. They will never hear me tell them that they are not cut out to do mathematics. I will not call bad mathematics good, but I will admit that good students can do bad mathematics. Moreover, I am willing to correct them as many times as I must while we move ahead. I am not going to blame some first-year-algebra teacher when my precalculus student squares $x + 3$ and gets $x^2 + 9$, nor will I assume that the student believes it forever to be true. I will correct the student and move on. Tomorrow I might have to correct the student again. I might even damage some self-esteem. But we will move on. And if that student learns some mathematics, shows some diligence, leaves my class with a 73 average, and goes on to make that same mistake in a college calculus class, then I would hope that the professor will not conclude that my student "cannot do algebra" and is therefore unworthy of taking a calculus course. The professor should correct the student again and move on. If the student should fail, which can happen for any number of reasons, then that student will have failed calculus under that professor, not first-year algebra post facto. And if that student had hoped to learn some calculus from that professor, then it will be a darned shame.

Not everyone will agree with the ideas in this article, but I hope that nobody will reject them as being impossible to implement. Assessment, however we define it, is only a means to an end: the learning of mathematics by all our students. If you suspect that your assessment is getting in the way of that basic goal, then I urge you to tame whatever beast it has become. The success of your students is in your hands—for your assessment is the instrument that defines that success.