WHAT GETS GRADED
IS WHAT GETS VALUED

Linda Wilson

A "MAVERICK" TEACHER
IN ACTION

An average day in algebra class found Ms. League asking her students about their homework assignments. A few responded with questions about particular textbook problems.

Self-assessment is a
central part of
authentic assessment.

Ms. League demonstrated solutions to the linear and quadratic systems of equations for about twenty minutes. She then assigned a problem for the students to try in their teaching pairs, which are the prearranged small groups or pairs that work together on various assignments. Members of these groups sat together to work. The problem was more abstract and complicated than those that the students had completed for homework. Ms. League suggested that they work on it by sketching a graph and making decisions from the graph.

As students worked on the problem, Ms. League gave help when needed. At one point she paused and interrupted their work by saying, "If you're having trouble with this, you need to write down, I need to study how to graph hyperbolas. Be sure to write that down."

Before she demonstrated the solution, she said, "When I was walking around, I saw all levels of thinking. Some of you were struggling with the graph. Others had trouble with the procedures. Remember that the test is on Friday, and although some of you did okay working together today, remember that on the test you will have to work alone. Be sure you know how much you can do alone before the test on Friday."

Students continued working on problems in pairs until the end of class, when Ms. League gave them their homework assignment for the next day—six more problems from the textbook similar to the ones done in class. She added, "Your other assignment is to look at what you wrote down today about what you need to do, for example, 'I need to review parabolas.' So do that tonight also."

Two aspects of this classroom are especially interesting from an assessment perspective. The first is that Ms. League routinely observed and interviewed her students as they worked in their teaching pairs, and the second is that she also emphasized ways students could assess their own work. She learned what her students knew and could do in mathematics by interviewing and observing their work in pairs. She knew her students well and could discuss their strengths and weaknesses at length.

VALUING CONCEPTUAL KNOWLEDGE

As with so many good teachers, Ms. League held a vast store of informal knowledge about her students in her head. It was also central to her theory about teaching. She believed that knowledge her students individually was fundamental to her teaching. She viewed observing and interviewing her students as they worked as one of the more authentic aspects of her assessment method.

Linda Wilson teaches at the University of Delaware, Newark, DE 19716. She is interested in assessment issues in the mathematics classroom and is assistant project director of the NCTM's Assessment Standards for School Mathematics.

Mathematics Teacher, September 1994
Ms. League asked her students to remind themselves of their strengths and weaknesses as learners. She frequently asked them to write comments to themselves about the concepts or procedures that needed more work or practice. Indeed, the NCTM's Assessment Standards draft endorses self-evaluation as a central aspect of authentic assessment and advocates that students keep journals noting their progress or that teachers invite students to be active participants in assessing their learning.

Ms. League would often ask students to respond to higher-order-thinking questions that went beyond the procedural aspects of a lesson. In the first lesson on logarithms, for example, she asked students to ponder why a log of a negative number is not possible. She clearly stated that she expected students to write individual explanations. These higher-order explanations were sometimes to be finished in class and sometimes part of a regular homework assignment that included textbook problems. It was evident, from frequent use of this type of assignment, that Ms. League valued her students' conceptual knowledge, not solely their ability to complete algebraic procedures.

Ms. League used several authentic assessment practices, including observing, interviewing, and asking students to write about mathematical concepts and their knowledge of them. One would expect her students to understand that talking and writing about mathematics were just as valued as completing routine procedures or algorithms. On the contrary, however, when students were interviewed, this perception was not evident, as the following excerpt suggests.

I: Have you ever had to write anything in mathematics class? Write a paragraph or explain something?
S: Yes, a report. I did a report once. That was for extra credit, though.
I: For what class was that?
S: Algebra 1. That was a while ago.

I: So you never had a situation where a teacher said, "Here's a problem. Tell me what the answer is. Now explain how you did it."
S: Yes, I have had that on a couple of tests before, too.
I: In this class?
S: Well, not in this class.

Students avoided activities that did not count toward their grade.

AN IMPORTANT LESSON TO BE LEARNED

Why is it that Ms. League's students did not recognize that they routinely were given writing assignments? The answer appears to rest in her grading system. At the end of the quarter, students received traditional A, B, C, D, or F grades, based on quizzes, tests, and examinations taken during that quarter. Homework was checked occasionally and points were given for its completion. Homework checks, however, did not include the extra assignments designed to foster self-assessment or higher-order thinking. Thus the extra assignments did not count toward the grade.

The tests and quizzes written by Ms. League were based on textbook problems and were identical to those assigned for homework, with the numbers having been changed. The tests determined whether students could complete the procedures or algorithms they had been doing in class for the previous week or ten days. They did not include the higher-order-thinking or self-evaluation questions. Students had to work individually and were not allowed to collaborate on the tests and quizzes.

The techniques described earlier as components of an authentic assessment system were not part of the grading system. Students, for the most part, avoided any activities that did not count toward their grade. On any given day, approximately half of them did not complete their homework. When Ms. League asked them to complete writing assignments in class or for homework, most of them ignored her. They did not do the self-evaluations, nor did they answer the higher-order-thinking questions. In fact, they did not recognize that writing was part of their algebra work.

The students, like most other high school students, were savvy about budgeting their energies. Only those tasks or activities that were reviewed and then recorded in the grade book received any effort or recognition. They paid little attention to any assessment activities other than quizzes, tests, examinations, and an occasional homework assignment. Because these activities were graded, they were valued by the students.

In the students' eyes, what counted as mathematical knowledge was the correct solution of decontextualized problems, such as a system of equations. Ms. League had a different notion of what counted as doing mathematics, which is illustrated by her higher-order questions and her exhortations to students to think, write, and work collaboratively. The nontraditional activities used reasoning, reflecting, and communicating more than routine procedures. But since the nontraditional activities were not graded, they were not valued by students.

An important lesson can be learned from this story. What do our students perceive as being of value in our classrooms? Many high school mathematics teachers are convinced, as Ms. League was, that using multiple sources of information about their students can lead to more valid inferences about what they know. Translating those beliefs into grading practices, though exceedingly more complicated, must be done or all our good efforts may be for naught. Our students know only too well to value only those items that get graded.
Grade Assignment Based on Progressive Improvement

The NCTM's Curriculum and Evaluation Standards for School Mathematics states, "Evaluation is a tool for implementing the Standards and effecting change systematically" (1989, 189). Tests are one facet of evaluation, and we maintain that mathematics classes are strongly affected by the way in which test scores are used to generate final course grades. In the traditional secondary school mathematics class, current grading practices tend to drive instruction by putting constraints on specific course content and its organization. In turn, content and its organization affect testing and therefore grading. The interaction of these factors is an aspect of assessment that is not specifically discussed by the NCTM's evaluation standards. The purpose of this article is to examine the impact of grading on mathematics instruction and on the implementation of the curriculum and evaluation standards.

Students' performance in today's secondary school mathematics curriculum is measured to a large extent by unit examinations administered on a regular basis. The numerical grades earned on these tests are averaged to help determine a final course grade. This article examines the logical consequences of this process of test-score averaging on the arrangement of course material, the nature of the tests themselves, and the learning emphasized. We believe that in each area, these consequences are negative. Furthermore, we believe that in line with the evaluation standards, test scores can be used to assign course grades in an objective and valid way that does not rely on averaging. An example of this type of grading will be described.

CONSEQUENCES OF AVERAGING

By the very nature of averages, examination scores earned early in the course are given equal weight with examination scores generated near the end. Thus, what students do not know early on in the course is given numerically equal weight with what they do not know at the end. Because early failures cannot be erased by subsequent learning, averaging also implies that specific learning must occur that we can test with a valid test after only a few weeks of the course.

Consider two examinations in a mathematics class, one given early on and the other given near the end of the term. Logically, the only way that they can be considered equal in assessing learning is if this learning is not regarded as cumulative. That is, each examination is truly a "unit" examination and the unit studied has no cumulative importance if it contributes nothing to subsequent concept development and if the procedures and concepts learned never reappear. These assumptions are false. The objections to averaging of test scores apply doubly to averaging of homework scores. Yet mathematics teachers continue to assign final course grades on the basis of averaged scores.

Grade averaging has an impact on the arrangement of course material and the types of tests that are used. Material tends to be compartmentalized into small, discrete "digestible chunks," with related short examination problems test specific and isolated skills. This approach allows teachers to demand immediate mastery of objectives with an accompanying 90-80-70-60 percent performance scale, yielding the grades they want to give.

Compartmentalizing course material sever connections between one topic and another. Concepts, which by definition must be abstracted from numerous examples in various contexts, are neglected in instruction because they cannot be mastered in small chunks at the 90-80-70-60 percent level during the short intervals between tests. Averaging of unit test scores forces teachers to compromise and write tests according to what the students can learn between tests, not according to what teachers really want the students to know.

For example, few students will fully grasp the difference between a function, f, and its image, f(x), immediately after functions are introduced.

Warren Esty teaches at Montana State University, Bozeman, MT 59717. He does research in applied probability and is working on his second freshman-level textbook. Anne Teppo is a mathematics education consultant, 1611 Willow Way, Bozeman, MT 59715. She is involved in qualitative research investigating the development of mathematical concepts within the classroom environment.
The distinction between the rule itself and the number created by applying the rule to \(x\) is subtle but extremely important. Calculus students need to know what \(f(x + h)\) is when \(f(x) = x^2\); yet many do not distinguish it from \(f(x) + h\) because they have not learned to isolate the rule, \(f'(x) = 2x\), from the image, \(x^2\). After the term function is introduced in algebra, a unit test is, of course, given that explores some aspects of particular functions. But the concept of function will not have been fully developed through instruction and certainly not yet fully understood by the students. Thus the teacher can ask questions only at a low level of comprehension if he or she is designing a unit test that students can pass. Success on such a test, although it improves a student’s chances for an acceptable final course grade, gives little indication of the student’s progress toward understanding the essential concept of function.

The evaluation standards recognize that tests “are one way of communicating what is important for students to know” (p. 189). Because students will not do well on conceptual questions, such as the distinction between \(f\) and \(f(x)\), such questions are usually not asked. Because conceptual questions are not asked, students fail to recognize the importance of concepts. Then whatever misconceptions the students develop on their own exist.

The pernicious effects of averaging carry through to higher-level courses.

Averaging and compartmentalizing reinforce one another. Without compartmentalizing averaging is inappropriate. With compartmentalizing averaging works. With averaging compartmentalizing seems necessary to justify the grading system. Therefore these two have settled in together. The compartmentalizing-averaging method of teaching and evaluating mathematics has come to be taken for granted. The vast majority of textbooks and courses depend on it. Teachers use it and it works—for material organized in compartments. But this type of organization is precisely what must be changed if mathematics courses are seriously to teach mathematical connections.

We need to assess concepts to emphasize their importance.

The evaluation standards recognize that “students’ grasp of mathematical concepts develops over time. Many concepts introduced in the early grades are later extended and studied in greater depth” (pp. 223–24). Clearly this type of development over time must happen throughout individual classes as well as from elementary to secondary grades. If we recognize that learning mathematics is cumulative and expect students to make connections among procedures and develop concepts over time, then we propose that it is acceptable—and completely expected—for students to be unable to demonstrate concept mastery until late in a course. The current practice of test-score averaging does not recognize these factors of time and complexity. The evaluation standards specifically assert that assessing what students do not know should receive decreased attention (p. 191), but averaging penalizes students for what they don’t know during the progress of the course instead of rewarding them for what they do know at the end.

The evaluation standards call for us “to reassess the manner and methods by which we chart our students’ progress” (p. 192). In terms of assigning students’ grades, the evaluation standards assert that assessment should measure (a) how well the student has understood and integrated the material, (b) if the student can apply his or her learning in other contexts, and (c) if the student is prepared to proceed to the next grade or level (p. 200). These questions are best answered only at the end of a mathematics course. They imply that assessment for grade assignment should be based on long-term course goals. Using a scoring method other than averaging for assigning class grades can free existing instruction from its present constraints and make it possible to emphasize the learning of complex concepts and related multistage procedures.

**GRADING BASED ON PROGRESSIVE IMPROVEMENT**

We have developed and implemented a grading system based on progressive improvement that measures the types of student progress described by the evaluation standards’ grading criteria. It is used for a ten-week course on the language and structure of mathematics offered to nonmathematically oriented students at Montana State University (Esty and Teppe 1991).

At the beginning of this class, students are informed that although they will be examined with diagnostic quizzes and tests throughout the course, only their performance in the final weeks of the class will be counted for assigning grades. They are told that what they learn will be cumulative and that at the end of the course they will be held responsible for all the material.

**RECOMMENDATIONS OF THE CURRICULUM AND EVALUATION STANDARDS**

The curriculum standards call for “a shift in emphasis from a curriculum dominated by memorization of isolated facts and procedures and by proficiency with paper-and-pencil skills to one that emphasizes conceptual understandings, multiple representations and connections, mathematical modeling, and mathematical problem solving” (p. 124). The averaging method currently employed in many mathematics classrooms inhibits the implementation of these recommendations.
Daily homework is checked for accuracy but not scored. Early quizzes and examinations are scored but not counted toward the final grade. Class participation is required and supplies useful feedback for both the instructor and student. Thus, assessment during the first two-thirds of the class is used to inform students and instructor of individual progress, not to generate course grades, that is, assessment is diagnostic and furnishes instructional feedback.

Students' final grades are not penalized by averaging in test scores that measure incorrect or incomplete understanding that occurs in the early stages of the course. Course grades are assigned on the basis of knowledge and skills displayed in the final few weeks of the class using quizzes and cumulative examinations, which are written tests, and, to a much lesser extent, class answers, which are spoken—yielding quite different information.

Because this approach is so nontraditional, we have had to deal with students' fears that we don't mean it and that in the end we really will average in some poor early score. We repeatedly reassure the students that we are looking for improvement and eventual mastery, not for a chance to take points off. For instance, quiz scores, instead of being rated on a 0-80-70-60 percent scale, are classified as "already at the C or A or B level" or "soon to be C or better."

Students' performance is expected to improve throughout the course. By omitting the requirement of immediate mastery, students can be held to a higher standard and posed more challenging problems. For example, on the second test administered halfway through the autumn 1990 course, a performance of 50 percent was judged by the instructor to be what was expected, at that time, of an eventual C student. The instructor's comment to the students on this "low" test score was, "You will have another chance soon to demonstrate that you have mastered the exam material." This recognition that examination scores were acceptable even if they did not fit a 90-80-70-60 percent scale is to be distinguished from curving, which is sometimes necessary under an averaging system according to which each score directly affects the student's final course grade.

Students' perceptions of the progressive grading system were obtained from interviews conducted as part of a qualitative study of the autumn 1990 course (Esty and Teppo 1991). Students remarked that this approach kept them working when they were not doing well on the diagnostic examinations and homework. "It keeps me feeling like I'm in the race," one student commented. Students knew they still had a chance to earn an acceptable final grade by continuing to improve throughout the course. As another student explained, "(You're given) a chance to make up the things you didn't understand after you've found a way to acquire the knowledge. You're rewarded for that, and I think it's really important."

This grading system acknowledges the reality that mathematics learning takes time. Instead of increasing anxiety early in the course for those who have not yet been able to grasp the material, this approach allows students the time actually required to put it all together and produces less negative reinforcement. Furthermore, less emphasis is placed on filing facts in short-term memory in cramming sessions just before unit examinations.

We have found that the students do not abuse this grading system—any more than they do any other grading system—by not working at the beginning of the course. As one interviewed student explained, "It isn't as though the teacher relieved us of the obligation or anything like that. He put all the students in a position where they could be confident in themselves." The students used the diagnostic quizzes to measure their performance as the course progressed. They understood that growth would occur and that they could realistically look forward to assimilating the material eventually. We have found that performance in the final weeks of the course accurately reflects this integration of knowledge throughout the course.

CONCLUSION

We maintain that the primary goal of instruction should be for students to master the material by the end of the course and that class grades should be assigned accordingly. We have found that the progressive-improvement grading system is a pedagogically effective way to assess and facilitate learning. By eliminating the artificial constraint on the material imposed by the test-score-averaging "instant mastery" requirement, the curriculum can be freed to include long-term learning of multi-stage procedures and broader mathematical concepts. Assessment changes are necessary if we are to implement the curriculum and evaluation standards.

REFERENCES


FREE CATALOG OF MATHEMATICS AND SCIENCE BOOKS

Over 400 books in math, physics, engineering and other sciences. Texts, monographs, classics many impossible to find elsewhere. Must $0.00 to $10.00. Write DOVER PUBLICATIONS, Dept. A308, 31 E. 2nd Street, Mineola, NY 11501.