GRADING

The Issue Is Not How but Why

By Alfie Kohn

Why are we concerned with evaluating how well students are doing? The question of motive, as opposed to method, can lead us to rethink basic tenets of teaching and learning and to evaluate what students have done in a manner more consistent with our ultimate educational objectives. But not all approaches to the topic result in this sort of thoughtful reflection. In fact, approaches to assessment may be classified according to their depth of analysis and willingness to question fundamental assumptions about how and why we grade. Consider three possible levels of inquiry:

Level 1. These are the most superficial concerns, those limited to the practical issue of how to grade students' work. Here we find articles and books offering elaborate formulas for scoring assignments, computing points, and allocating final grades -- thereby taking for granted that what students do must receive some grades and, by extension, that students ought to be avidly concerned about the ones they will get.

Level 2. Here educators call the above premises into question, asking whether traditional grading is really necessary or useful for assessing students' performance. Alternative assessments, often designated as "authentic," belong in this category. The idea here is to provide a richer, deeper description of students' achievement. (Portfolios of students' work are sometimes commended to us in this context, but when a portfolio is used merely as a means of arriving at a traditional grade, it might more accurately be grouped under Level 1.)

Level 3. Rather than challenging grades alone, discussions at this level challenge the whole enterprise of assessment -- and specifically why we are evaluating students as opposed to how we are doing so. No matter how elaborate or carefully designed an assessment strategy may be, the result will not be constructive if our reason for wanting to know how students are doing is itself objectionable.

Grading Rationale I: Sorting

One reason for evaluating students is to be able to label them on the basis of their performance and thus to sort them like so many potatoes. Sorting, in turn, has been criticized at each of the three levels, but for very different reasons. At Level 1, the concern is merely that we are not correctly dumping individuals into the right piles. The major problem with our high schools and colleges, the argument goes, is that they don't keep enough students off the Excellent pile. (These critics don't put it quite this way, of course; they talk about "grade inflation.") Interestingly, most studies suggest that student performance does not improve when instructors grade more stringently and, conversely, that making it relatively easy to get a good grade does not lead students to do inferior work -- even when performance is defined as the number of facts retained temporarily as measured by multiple-choice exams (Vasta and Sarmiento 1979, Abrami et al. 1980).

At Level 2, questions are raised about whether grades are reliable enough to allow students

to be sorted effectively. Indeed, studies show that any particular teacher may well give different grades to a single piece of work submitted at two different times. Naturally the variation is even greater when the work is evaluated by more than one teacher (Kirschenbaum et al. 1971). What grades offer is spurious precision, a subjective rating masquerading as an objective assessment.

From the perspective of Level 3, this criticism is far too tame. The trouble is not that we are sorting students badly -- a problem that logically should be addressed by trying to do it better. The trouble is that we are sorting them at all. Are we doing so in order to segregate students by ability and teach them separately? The harms of this practice have been well established (Oakes 1985). Are we turning schools into "bargain-basement personnel screening agencies for business" (Campbell 1974, p. 145)? Whatever use we make of sorting, the process itself is very different from -- and often incompatible with -- the goal of helping students to learn.

Grading Rationale II: Motivation

A second rationale for grading -- and indeed, one of the major motives behind assessment in general -- is to motivate students to work harder so they will receive a favorable evaluation. Unfortunately, this rationale is just as problematic as sorting. Indeed, given the extent to which A's and F's function as rewards and punishments rather than as useful feedback, grades are counterproductive regardless of whether they are intentionally used for this purpose. The trouble lies with the implicit assumption that there exists a single entity called "motivation" that students have to a greater or lesser degree. In reality, a critical and qualitative difference exists between intrinsic and extrinsic motivation -- between an interest in what one is learning for its own sake, and a mindset in which learning is viewed as a means to an end, the end being to escape a punishment or snag a reward. Not only are these two orientations distinct, but they also often pull in opposite directions.

Scores of studies in social psychology and related fields have demonstrated that extrinsic motivators frequently undermine intrinsic motivation. This may not be particularly surprising in the case of sticks, but it is no less true of carrots. People who are promised rewards for doing something tend to lose interest in whatever they had to do to obtain the reward. Studies also show that, contrary to the conventional wisdom in our society, people who have been led to think about what they will receive for engaging in a task (or for doing it well) are apt to do lower quality work than those who are not expecting to get anything at all.

These findings are consistent across a variety of subject populations, rewards, and tasks, with the most destructive effects occurring in activities that require creativity or higher-order thinking. That this effect is produced by the extrinsic motivators known as grades has been documented with students of different ages and from different cultures. Yet the findings are rarely cited by educators.

Studies have shown that the more students are induced to think about what they will get on an assignment, the more their desire to learn evaporates, and, ironically, the less well they do. Consider these findings:

* On tasks requiring varying degrees of creativity, Israeli educational psychologist Ruth Butler has repeatedly found that students perform less well and are less interested in what they are doing when being graded than when they are encouraged to focus on the task itself (Butler and Nissan 1986; Butler 1987, 1988).

* Even in the case of rote learning, students are more apt to forget what they have learned after a week or so -- and are less apt to find it interesting -- if they are initially advised that

they will be graded on their performance (Grolnick and Ryan 1987).

* When Japanese students were told that a history test would count toward their final grade, they were less interested in the subject -- and less likely to prefer tackling difficult questions than those who were told the test was just for monitoring their progress (Kage 1991).

* Children told that they would be graded on their solution of anagrams chose easier ones to work on -- and seemed to take less pleasure from solving them -- than children who were not being graded (Harter 1978).

As an article in the Journal of Educational Psychology concluded, "Grades may encourage an emphasis on quantitative aspects of learning, depress creativity, foster fear of failure, and undermine interest" (Butler and Nissan 1986, p. 215). This is a particularly ironic result if the rationale for evaluating students in the first place is to encourage them to perform better.

Grading Rationale III: Feedback

Some educators insist that their purpose in evaluating students is neither to sort them nor to motivate them, but simply to provide feedback so they can learn more effectively tomorrow than they did today. From a Level 2 perspective, this is an entirely legitimate goal -- and grades are an entirely inadequate means of reaching it. There is nothing wrong with helping students to internalize and work toward meeting high standards, but that is most likely to happen when they "experience success and failure not as reward and punishment, but as information" (Bruner 1961, p. 26). Grades make it very difficult to do this. Besides, reducing someone's work to a letter or number simply is not helpful; a B+ on top of a paper tells a student nothing about what was impressive about that paper or how it could be improved.

But from Level 3 comes the following challenge: Why do we want students to improve? This question at first seems as simple and bland as baby food; only after a moment does it reveal a jalapeño kick: it leads us into disconcerting questions about the purpose of education itself.

Demand vs. Support

Eric Schaps (1993), who directs the Developmental Studies Center in Oakland, California, has emphasized "a single powerful distinction: focusing on what students ought to be able to do, that is, what we will demand of them -- as contrasted with focusing on what we can do to support students' development and help them learn." For lack of better labels, let us call these the "demand" and "support" models.

In the demand model, students are workers who are obligated to do a better job. Blame is leveled by saying students "chose" not to study or "earned" a certain grade -- conveniently removing all responsibility from educators and deflecting attention from the curriculum and the context in which it is taught. In their evaluations, teachers report whether students did what they were supposed to do. This mind-set often lurks behind even relatively enlightened programs that emphasize performance assessment and -- a common buzzword these days -- outcomes. (It also manifests itself in the view of education as an investment, a way of preparing children to become future workers.)

The support model, by contrast, helps children take part in an "adventure in ideas" (Nicholls and Hazzard 1993), guiding and stimulating their natural inclination to explore what is unfamiliar; to construct meaning; to develop a competence with and a passion for playing with words, numbers, and ideas. This approach meshes with what is sometimes called "learner-centered" learning, in which the point is to help students act on their desire to make sense of the world. In this context, student evaluation is, in part, a way of determining how

effective we have been as educators. In sum, improvement is not something we require of students so much as something that follows when we provide them with engaging tasks and a supportive environment.

Supportive Assessment

Here are five principles of assessment that follow from this support model:

1. Assessment of any kind should not be overdone. Getting students to become preoccupied with how they are doing can undermine their interest in what they are doing. An excessive concern with performance can erode curiosity -- and, paradoxically, reduce the quality of performance. Performance-obsessed students also tend to avoid difficult tasks so they can escape a negative evaluation.

2. The best evidence we have of whether we are succeeding as educators comes from observing children's behavior rather than from test scores or grades. It comes from watching to see whether they continue arguing animatedly about an issue raised in class after the class is over, whether they come home chattering about something they discovered in school, whether they read on their own time. Where interest is sparked, skills are usually acquired. Of course, interest is difficult to quantify, but the solution is not to return to more conventional measuring methods; it is to acknowledge the limits of measurement.

3. We must transform schools into safe, caring communities. This is critical for helping students to become good learners and good people, but it is also relevant to assessment. Only in a safe place, where there is no fear of humiliation and punitive judgment, will students admit to being confused about what they have read and feel free to acknowledge their mistakes. Only by being able to ask for help will they be likely to improve.

Ironically, the climate created by an emphasis on grades, standardized testing, coercive mechanisms such as pop quizzes and compulsory recitation, and pressure on teachers to cover a prescribed curriculum makes it more difficult to know how well students understand -- and thus to help them along.

4. Any responsible conversation about assessment must attend to the quality of the curriculum. The easy question is whether a student has learned something; the far more important -- and unsettling -- question is whether the student has been given something worth learning. (The answer to the latter question is almost certainly no if the need to evaluate students has determined curriculum content.) Research corroborates what thoughtful teachers know from experience: when students have interesting things to do, artificial inducements to boost achievement are unnecessary (Moeller and Reschke 1993).

5. Students must be invited to participate in determining the criteria by which their work will be judged, and then play a role in weighing their work against those criteria. Indeed, they should help make decisions about as many elements of their learning as possible (Kohn 1993). This achieves several things: It gives them more control over their education, makes evaluation feel less punitive, and provides an important learning experience in itself. If there is a movement away from grades, teachers should explain the rationale and solicit students' suggestions for what to do instead and how to manage the transitional period. That transition may be bumpy and slow, but the chance to engage in personal and collective reflection about these issues will be important in its own right.

And If You Must Grade ...

Finally, while conventional grades persist, teachers and parents ought to do everything in

their power to help students forget about them. Here are some practical suggestions for reducing the salience.

* Refrain from giving a letter or number grade for individual assignments, even if you are compelled to give one at the end of the term. The data suggest that substantive comments should replace, not supplement, grades (Butler 1988). Make sure the effect of doing this is not to create suspense about what students are going to get on their report cards, which would defeat the whole purpose. Some older students may experience, especially at first, a sense of existential vertigo: a steady supply of grades has defined them. Offer to discuss privately with any such student the grade he or she would probably receive if report cards were handed out that day. With luck and skill, the requests for ratings will decrease as students come to be involved in what is being taught.

* Never grade students while they are still learning something and, even more important, do not reward them for their performance at that point. Studies suggest that rewards are most destructive when given for skills still being honed (Condry and Chambers 1978). If it is unclear whether students feel ready to demonstrate what they know, there is an easy way to find out: ask them.

* Never grade on a curve. The number of good grades should not be artificially limited so that one student's success makes another's less likely. Stipulating that only a few individuals can get top marks regardless of how well everyone does is egregiously unfair on its face. It also undermines collaboration and community. Of course, grades of any kind, even when they are not curved to create artificial scarcity -- or deliberately publicized -- tend to foster comparison and competition, an emphasis on relative standing. This is not only destructive to students' self-esteem and relationships but also counterproductive with respect to the quality of learning (Kohn 1992). As one book on the subject puts it: "It is not a symbol of rigor to have grades fall into a 'normal' distribution; rather, it is a symbol of failure: failure to teach well, to test well, and to have any influence at all on the intellectual lives of students" (Milton et al. 1986, p. 225).

* Never give a separate grade for effort. When students seem to be indifferent to what they are being asked to learn, educators sometimes respond with the very strategy that precipitated the problem in the first place: grading students' efforts to coerce them to try harder. The fatal paradox is that while coercion can sometimes elicit resentful obedience, it can never create desire. A low grade for effort is more likely to be read as "You're a failure even at trying." On the other hand, a high grade for effort combined with a low grade for achievement says, "You're just too dumb to succeed." Most of all, rewarding or punishing children's efforts allows educators to ignore the possibility that the curriculum or learning environment may have something to do with students' lack of enthusiasm.

NOTES

Abrami, P. C., W. J. Dickens, R. P. Perry, and L. Leventhal. (1980). "Do Teacher Standards for Assigning Grades Affect Student Evaluations of Instruction?" Journal of Educational Psychology 72: 107-118.

Bruner, J. S. (196 1). "The Act of Discovery." Harvard Educational Review 31: 21-32.

Butler, R. (1987). "Task-Involving and Ego-Involving Properties of Evaluation." Journal of Educational Psychology 79: 474-482.

Butler, R. (1988) Enhancing and Undermining Intrinsic Motivation." British Journal of Educational Psychology 58 (1988): 1-14.

Butler, R., and M. Nissan. (1986). "Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance." Journal of' Educational Psychology 78: 210-216.

Campbell, D. N. (October 1974) "On Being Number One: Competition in Education." Phi Delta Kappan: 143-146.

Condry, J., and J. Chambers. (1978). "Intrinsic Motivation and the Process of Learning." In The Hidden Costs of Rewards: New Perspectives on the Psychology of Human Motivation, edited by M. R. Lepper and D. Greene. Hillsdale, N.J.: Lawrence Erlbaum.

Grolnick, W. S., and R. M. Ryan. (1987). "Autonomy in Children's Learning: An Experimental and Individual Difference Investigation." Journal of Personality and Social Psychology 52: 890-898.

Harter, S. (1978). "Pleasure Derived from Challenge and the Effects of Receiving Grades on Children's Difficulty Level Choices." Child Development 49: 788-799.

Kage, M. (1991). "The Effects of Evaluation on Intrinsic Motivation." Paper presented at the meeting of the Japan Association of Educational Psychology, Joetsu, Japan.

Kirschenbaum, H., R. W. Napier, and S. B. Simon. (1971). Wad-Ja-Get?: The Grading Game in American Education. New York: Hart.

Kohn, A. (1992). No Contest: The Case Against Competition. Rev. ed. Boston: Houghton Mifflin.

Kohn, A. (September 1993). "Choices for Children: Why and How to Let Students Decide." Phi Delta Kappan: 8-20.

Milton, O., H. R. Pollio, and J. A. Eison. (1986). Making Sense of College Grades. San Francisco: Jossey-Bass.

Moeller, A. J., and C. Reschke. (1993). "A Second Look at Grading and Classroom Performance." Modern Language Journal 77: 163-169.

Nicholls, J. G., and S. P. Hazzard. (1993). Education as Adventure: Lessons from the Second Grade. New York: Teachers College Press.

Oakes, J. (1985). Keeping Track: How Schools Structure Inequality. New Haven: Yale University Press.

Schaps, E. (October 1993). Personal communication.

Vasta, R., and R. F. Sarmiento. (1979). "Liberal Grading Improves Evaluations But Not Performance," Journal of Educational Psychology 71: 207-211.

Back to Top   Print Article   Close this Window

www.alfiekohn.org -- © Alfie Kohn