Making the Grade: What Benefits Students?
Thomas R. Guskey
Although the debate over grading and reporting practices continues, today we know which practices benefit students and encourage learning.

Charged with leading a committee that would revise his school's grading and reporting system, Warren Middleton described his work this way:

The Committee On Grading was called upon to study grading procedures. At first, the task of investigating the literature seemed to be a rather hopeless one. What a mass and a mess it all was! Could order be brought out of such chaos? Could points of agreement among American educators concerning the perplexing grading problem actually be discovered? It was with considerable misgiving and trepidation that the work was finally begun.

Few educators today would consider the difficulties encountered by Middleton and his colleagues to be particularly surprising. In fact, most probably would sympathize with his lament. What they might find surprising, however, is that this report from the Committee on Grading was published in 1933! The issues of grading and reporting on student learning have perplexed educators for the better part of this century. Yet despite all the debate and the multitude of studies, coming up with prescriptions for best practice seems as challenging today as it was for Middleton and his colleagues more than 60 years ago.

Points of Agreement
Although the debate over grading and reporting continues, today we know better which practices benefit students and encourage learning. Given the multitude of studies—and their often incongruous results—researchers do appear to agree on the following points:

1. *Grading and reporting aren't essential to instruction.* Teachers don't need grades or reporting forms to teach well. Further, students don't need them to learn (Frisbie and Waltman 1992). Teachers do need to check regularly on how students are doing, what they've learned, and what problems or difficulties they've experienced. But grading and reporting are different from checking; they involve judging the adequacy of students' performance at a specific time. Typically, teachers use checking to diagnose and prescribe and use grading to evaluate and describe (Bloom et al. 1981).

   When teachers do both checking and grading, they become advocates as well as judges—roles that aren't necessarily compatible (Bishop 1992). Finding a meaningful compromise between these dual roles makes many teachers uncomfortable, especially those with a child-centered orientation (Barnes 1985).

2. *No one method of grading and reporting serves all purposes well.* Grading enables teachers to communicate the achievements of students to parents and others, provide incentives to learn, and

provide information that students can use for self-evaluation. In addition, schools use grades to identify or group students for particular educational paths or programs and to evaluate a program's effectiveness (Feldmesser 1971, Frisbie and Waltman 1992). Unfortunately, many schools attempt to address all of these purposes with a single method and end up achieving none very well (Austin and McCann 1992).

Letter grades, for example, briefly describe learning progress and give some idea of its adequacy (Payne 1974). Their use, however, requires abstracting a great deal of information into a single symbol (Stiggins 1994). In addition, the cut-off between grade categories is always arbitrary and difficult to justify. If scores for a grade of *B* range from 80 to 89, students at both ends of that range receive the same grade, even though their scores differ by nine points. But the student with a score of 79—a one-point difference—receives a grade of *C*.

The more detailed methods also have their drawbacks. Narratives and checklists of learning outcomes offer specific information for documenting progress, but good narratives take time to prepare, and—not surprisingly—as teachers complete more narratives, their comments become increasingly standardized. From the parents' standpoint, checklists of learning outcomes often appear too complicated to understand. In addition, checklists seldom communicate the appropriateness of students' progress in relation to expectations for their level (Afflerbach and Sammons 1991).

Because one method won't adequately serve all purposes, schools must identify their primary purpose for grading and select or develop the most appropriate approach (Cangelosi 1990). This process often involves the difficult task of seeking consensus among several constituencies.

3. *Regardless of the method used, grading and reporting remain inherently subjective.* In fact, the more detailed the reporting method and the more analytic the process, the more likely subjectivity will influence results (Ornstein 1994). That's why, for example, holistic scoring procedures tend to have greater reliability than analytic procedures.

Subjectivity in this process, however, isn't always bad. Because teachers know their students, understand various dimensions of students' work, and have clear notions of the progress made, their subjective perceptions may yield very accurate descriptions of what students have learned (Brookhart 1993, O'Donnell and Woolfolk 1991).

When subjectivity translates into bias, however, negative consequences can result. Teachers' perceptions of students' behavior can significantly influence their judgments of scholastic performance (Hills 1991). Students with behavior problems often have no chance to receive a high grade because their infractions overshadow their performance. These effects are especially pronounced in judgments of boys (Bennett et al. 1993). Even the neatness of students' handwriting can significantly affect a teacher's judgment (Sweedler-Brown 1992)

Training programs can help teachers identify and reduce these negative effects and lead to greater consistency in judgments (Afflerbach and Sammons 1991). Unfortunately, few teachers receive adequate training in grading or reporting as part of their preservice experiences (Boothroyd and McMorris 1992). Also, few school districts provide adequate guidance to ensure consistency in teachers' grading or reporting practices (Austin and McCann 1992).

4. *Grades have some value as rewards, but no value as punishments.* Although educators would undoubtedly prefer that motivation to learn be entirely intrinsic, the existence of grades and other reporting methods are important factors in determining how much effort students put forth (Chastain 1990, Ebel 1979). Most students view high grades as positive recognition of their success, and some work hard to avoid the consequences of low grades (Feldmesser 1971).

At the same time, no studies support the use of low grades as punishments. Instead of prompting greater effort, low grades usually cause students to withdraw from learning. To protect their self-image, many students regard the low grade as irrelevant and meaningless. Other students may blame themselves for the low mark, but feel helpless to improve (Selby and Murphy 1992). Sadly, some teachers consider grades or reporting forms their "weapon of last resort." In their view, students who don't comply with requests suffer the consequences of the greatest punishment a teacher can bestow: a failing grade. Such practices have no educational value and, in the long run, adversely affect students, teachers, and the relationship they share. Rather than attempting to punish students with a low mark, teachers can better motivate students by regarding their work as incomplete and requiring additional effort.

5. *Grading and reporting should always be done in reference to learning criteria, never on the curve.* Using the normal probability curve as a basis for assigning grades typically yields greater consistency in grade distributions from one teacher to the next. The practice, however, is detrimental to teaching and learning.

Grading on the curve pits students against one another in a competition for the few rewards (high grades) distributed by the teacher. Under these conditions, students readily see that helping others will threaten their own chances for success (Johnson et al. 1979, Johnson et al. 1980). Learning becomes a game of winners and losers—with most students falling into the latter category (Johnson and Johnson 1989). In addition, modern research has shown that the seemingly direct relationship between aptitude or intelligence and school achievement depends upon instructional conditions, not a probability curve.

When the instructional quality is high and well matched to students' learning needs, the magnitude of this relationship diminishes drastically and approaches zero (Bloom 1976). Moreover, the fairness and equity of grading on the curve is a myth.

Learning Criteria
When grading and reporting relate to learning criteria, teachers have a clearer picture of what students have learned. Students and teachers alike generally prefer this approach because it seems fairer (Kovas 1993). The types of learning criteria usually used for grading and reporting fall into three categories:

- *Product criteria* are favored by advocates of performance-based approaches to teaching and learning. These educators believe grading and reporting should communicate a summative evaluation of student achievement (Cangelosi 1990). In other words, they focus on what students know and are able to do at that time. Teachers who use product criteria often base their grades or reports exclusively on final examination scores, overall assessments, or other culminating demonstrations of learning.

- *Process criteria* are emphasized by educators who believe product criteria don't provide a complete picture of student learning. From their perspective, grading and reporting should reflect not just the final results but also *how* students got there. Teachers who consider effort or work habits when reporting on student learning are using process criteria. So are teachers who take into consideration classroom quizzes, homework, class participation, or attendance.

- *Progress criteria,* often referred to as "improvement scoring" and "learning gain," consider how much students have gained from their learning experiences. Teachers who use progress criteria look at *how far* students have come rather than where they are. As a result, scoring criteria may become highly individualized.

Teachers who base their grading and reporting procedures on learning criteria typically use some combination of the three types (Frary et al. 1993; Nava and Loyd 1992; Stiggins et al. 1989). Most researchers and measurement specialists, on the other hand, recommend using product criteria exclusively. They point out that the more process and progress criteria come into play, the more subjective and biased grades become (Ornstein 1994). How can a teacher know, for example, how difficult a task was for students or how hard they worked to complete it? If these criteria are included at all, most experts recommend they be reported separately (Stiggins 1994).

Practical Guidelines

Despite years of research, there's no evidence to indicate that one grading or reporting method works best under all conditions, in all circumstances. But in developing practices that seek to be fair, equitable, and useful to students, parents, and teachers, educators can rely on two guidelines:

- *Provide accurate and understandable descriptions of learning.* Regardless of the method or form used, grading and reporting should communicate effectively what students have learned, what they can do, and whether their learning status is in line with expectations for that level. More than an exercise in quantifying achievement, grading and reporting must be seen as a challenge in clear thinking and effective communication (Stiggins 1994).

- *Use grading and reporting methods to enhance, not hinder, teaching and learning.* A clear, easily understood reporting form facilitates communication between teachers and parents. When both parties speak the same language, joint efforts to help students are likely to succeed. But developing such an equitable and understandable system will require the elimination of long-time practices such as averaging and assigning a zero to work that's late, missed, or neglected.

Averaging falls far short of providing an accurate description of what students have learned. For example, students often say, "I have to get a *B* on the final to pass this course." Such a comment illustrates the inappropriateness of averaging. If a final examination is truly comprehensive and students' scores accurately reflect what they've learned, why should a *B* level of performance translate to a *D* for the course grade?

Any single measure of learning can be unreliable. Consequently, most researchers recommend using several indicators in determining students' grades or marks—and most teachers concur (Natriello 1987). Nevertheless, the key question remains, "What information provides the most accurate depiction of students' learning at this time?" In nearly all cases, the answer is "the most current information." If students demonstrate that past assessment information doesn't accurately reflect their learning, new information must take its place. By continuing to rely on past assessment data, the grades can be misleading about a student's learning (Stiggins 1994).

Similarly, assigning a score of zero to work that is late, missed, or neglected doesn't accurately depict learning. Is the teacher certain the student has learned absolutely nothing, or is the zero assigned to punish students for not displaying appropriate responsibility (Canady and Hotchkiss 1989, Stiggins and Duke 1991)?

Further, a zero has a profound effect when combined with the practice of averaging. Students who receive a single zero have little chance of success because such an extreme score skews the average. That is why, for example, Olympic events such as gymnastics and ice skating eliminate the highest and lowest scores; otherwise, one judge could control the entire competition simply by giving extreme scores. An alternative is to use the median score rather than the average (Wright 1994), but use of the most current information remains the most defensible option.

Meeting the Challenge

The issues of grading and reporting on student learning continue to challenge educators today, just as they challenged Middleton and his colleagues in 1933. But today we know more than ever before about the complexities involved and how certain practices can influence teaching and learning. What do educators need to develop grading and reporting practices that provide quality information about student learning? Nothing less than clear thinking, careful planning, excellent communication skills, and an overriding concern for the well being of students. Combining these skills with our current knowledge on effective practice will surely result in more efficient and more effective reporting.

A Look Back at Grading Practices

Although student assessment has been a part of teaching and learning for centuries, grading is a relatively recent phenomenon. The ancient Greeks used assessments as formative, not evaluative, tools. Students demonstrated, usually orally, what they had learned, giving teachers a clear indication of which topics required more work or instruction.

In the United States, grading and reporting were virtually unknown before 1850. Back then, most schools grouped students of all ages and backgrounds together with one teacher. Few students went beyond the elementary education offered in these one-room schoolhouses. As the country grew—and as legislators passed compulsory attendance laws—the number and diversity of students increased. Schools began to group students in grades according to their age, and to try new ideas about curriculum and teaching methods. Here's a brief timeline of significant dates in the history of grading:

Late 1800s: Schools begin to issue progress evaluations. Teachers simply write down the skills that students have mastered; once students complete the requirements for one level, they can move to the next level.

Early 1900s: The number of public high schools in the United States increases dramatically. While elementary teachers continue using written descriptions to document student learning, high school teachers introduce percentages as a way to certify students' accomplishments in specific subject areas. Few educators question the gradual shift to percentage grading, which seems a natural by-product of the increased demands on high school teachers.

1912: Starch and Elliott publish a study that challenges percentage grades as reliable measures of student achievement. They base their findings on grades assigned to two papers written for a first-year English class in high school. Of the 142 teachers grading on a 0 to 100 scale, 15 percent give one paper a failing mark; 12 percent give the same paper a score of 90 or more. The other paper receives scores ranging from 50 to 97. Neatness, spelling, and punctuation influenced the scoring of many teachers, while others considered how well the paper communicated its message.

1913: Responding to critics—who argue that good writing is, by nature, a highly subjective judgment—Starch and Elliott repeat their study but use geometry papers. Even greater variations occur, with scores on one paper ranging from 28 to 95. Some teachers deducted points only for wrong answers, but others took neatness, form, and spelling into account.

1918: Teachers turn to grading scales with fewer and larger categories. One three-point scale, for example, uses the categories of Excellent, Average, and Poor. Another has five categories (Excellent, Good, Average, Poor, and Failing) with the corresponding letters of *A, B, C, D,* and *F* (Johnson 1918, Rugg 1918).

1930s: Grading on the curve becomes increasingly popular as educators seek to minimize the subjective nature of scoring. This method rank orders students according to some measure of their performance or

proficiency. The top percentage receives an *A*, the next percentage receives a *B*, and so on (Corey 1930). Some advocates (Davis 1930) even specify the precise percentage of students to be assigned each grade, such as 6-22-44-22-6.

Grading on the curve seems fair and equitable, given research suggesting that students' scores on tests of innate intelligence approximate a normal probability curve (Middleton 1933).

As the debate over grading and reporting intensifies, a number of schools abolish formal grades altogether (Chapman and Ashbaugh 1925) and return to using verbal descriptions of student achievement. Others advocate pass-fail systems that distinguish only between acceptable and failing work (Good 1937). Still others advocate a "mastery approach": Once students have mastered a skill or content, they move to other areas of study (Heck 1938, Hill 1935).

1958: Ellis Page investigates how student learning is affected by grades and teachers' comments. In a now classic study, 74 secondary school teachers administer a test, and assign a numerical score and letter grade of *A, B, C, D*, or *F* to each student's paper. Next, teachers randomly divide the tests into three groups. Papers in the first group receive only the numerical score and letter grade. The second group, in addition to the score and grade, receive these standard comments: *A—Excellent! B—Good work. Keep at it. C—Perhaps try to do still better? D—Let's bring this up. F—Let's raise this grade!* For the third group, teachers mark the score and letter grade, and write individualized comments.

Page evaluates the effects of the comments by considering students' scores on the next test they take. Results show that students in the second group achieved significantly higher scores than those who received only a score and grade. The students who received individualized comments did even better. Page concludes that grades can have a beneficial effect on student learning, but only when accompanied by specific or individualized comments from the teacher.

—Thomas R. Guskey

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

References

Afflerbach, P., and R. B. Sammons. (1991). "Report Cards in Literacy Evaluation: Teachers' Training, Practices, and Values." Paper presented at the annual meeting of the National Reading Conference, Palm Springs, Calif.

Austin, S., and R. McCann. (1992). "`Here's Another Arbitrary Grade for your Collection': A Statewide Study of Grading Policies." Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Barnes, S. (1985). "A Study of Classroom Pupil Evaluation: The Missing Link in Teacher Education."*Journal of Teacher Education* 36, 4: 46–49.

Bennett, R. E., R. L. Gottesman, D. A. Rock, and F. Cerullo. (1993). "Influence of Behavior Perceptions and Gender on Teachers' Judgments of Students' Academic Skill." *Journal of Educational Psychology*, 85: 347–356.

Bishop, J. H. (1992). "Why U.S. Students Need Incentives to Learn." *Educational Leadership* 49, 6: 15–18.

Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.

Bloom, B. S., G. F. Madaus, and J. T. Hastings. (1981).*Evaluation to Improve Learning*. New York: McGraw-Hill.

Boothroyd, R. A., and R. F. McMorris. (1992). "What Do Teachers Know About Testing and How Did They Find Out?" Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Brookhart, S. M. (1993). "Teachers' Grading Practices: Meaning and Values." *Journal of Educational Measurement* 30, 2: 123–142.

Canady, R. L., and P. R. Hotchkiss. (1989). "It's a Good Score! Just a Bad Grade." *Phi Delta Kappan* 71: 68–71.

Cangelosi, J. S. (1990). "Grading and Reporting Student Achievement." In *Designing Tests for Evaluating Student Achievement*, pp. 196–213. New York: Longman.

Chapman, H. B., and E. J. Ashbaugh. (October 7, 1925). "Report Cards in American Cities."*Educational Research Bulletin* 4: 289–310.

Chastain, K. (1990). Characteristics of Graded and Ungraded Compositions." *Modern Language Journal*, 74, 1: 10–14.

Corey, S. M. (1930). "Use of the Normal Curve as a Basis for Assigning Grades in Small Classes."*School and Society* 31: 514–516.

Davis, J. D. W. (1930). "Effect of the 6-22-44-22-6 Normal Curve System on Failures and Grade Values." *Journal of Educational Psychology* 22: 636–640.

Ebel, R. L. (1979). *Essentials of Educational Measurement* (3rd ed.). Englewood Cliffs, N.J.: Prentice Hall.

Feldmesser, R. A. (1971). "The Positive Functions of Grades." Paper presented at the annual meeting of the American Educational Research Association, New York.

Frary, R. B., L. H. Cross, and L. J. Weber. (1993). "Testing and Grading Practices and Opinions of Secondary Teachers of Academic Subjects: Implications for Instruction in Measurement."*Educational Measurement: Issues and Practices* 12, 3: 23–30.

Frisbie, D. A., and K. K. Waltman. (1992). "Developing a Personal Grading Plan." *Educational Measurement: Issues and Practices* 11, 3: 35–42.

Good, W. (1937). "Should Grades Be Abolished?"*Education Digest* 2, 4: 7–9.

Heck, A. O. (1938). "Contributions of Research to Classification, Promotion, Marking and Certification." Reported in*The Science Movement in Education (Part II), Twenty-Seventh Yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press.

Hill, G. E. (1935). "The Report Card in Present Practice." *Education Methods* 15, 3: 115–131.

Hills, J. R. (1991). "Apathy Concerning Grading and Testing." *Phi Delta Kappan* 72, 2: 540–545.

Johnson, D. W., and R. T. Johnson. (1989). *Cooperation and Competition: Theory and Research*. Endina, Minn.: Interaction.

Johnson, D. W., L. Skon, and R. T. Johnson. (1980). "Effects of Cooperative, Competitive, and Individualistic Conditions on Children's Problem-Solving Performance." *American Educational Research Journal* 17, 1: 83–93.

Johnson, R. H. (1918). "Educational Research and Statistics: The Coefficient Marking System."*School and Society* 7, 181: 714–116.

Johnson, R. T., D. W. Johnson, and M. Tauer. (1979). "The Effects of Cooperative, Competitive, and Individualistic Goal Structures on Students' Attitudes and Achievement." *Journal of Psychology* 102: 191–198.

Kovas, M. A. (1993). "Make Your Grading Motivating: Keys to Performance-Based Evaluation." *Quill and Scroll* 68, 1: 10–11.

Middleton, W. (1933). "Some General Trends in Grading Procedure." *Education* 54, 1: 5–10.

Natriello, G. (1987). "The Impact of Evaluation Processes On Students." *Educational Psychologists* 22: 155–175.

Nava, F. J. G., and B. H. Loyd. (1992). "An Investigation of Achievement and Nonachievement Criteria in Elementary and Secondary School Grading." Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

O'Donnell, A., and A. E. Woolfolk. (1991). "Elementary and Secondary Teachers' Beliefs About Testing and Grading." Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Ornstein, A. C. (1994). "Grading Practices and Policies: An Overview and Some Suggestions." *NASSP Bulletin* 78, 559: 55–64.

Page, E. B. (1958). "Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School Motivation."*Journal of Educational Psychology* 49: 173–181.

Payne, D. A. (1974). *The Assessment of Learning*. Lexington, Mass.: Heath.

Rugg, H. O. (1918). "Teachers' Marks and the Reconstruction of the Marking System." *Elementary School Journal* 18, 9: 701–719.

Selby, D., and S. Murphy. (1992). "Graded or Degraded: Perceptions of Letter-Grading for Mainstreamed Learning-Disabled Students." *British Columbia Journal of Special Education* 16, 1: 92–104.

Starch, D., and E. C. Elliott. (1912). "Reliability of the Grading of High School Work in English." *School Review* 20: 442–457.

Starch, D., and E. C. Elliott. (1913). "Reliability of the Grading of High School Work in Mathematics."*School Review* 21: 254–259.

Stewart, L. G., and M. A. White. (1976). "Teacher Comments, Letter Grades, and Student Performance." *Journal of Educational Psychology* 68, 4: 488–500.

Stiggins, R. J. (1994). "Communicating with Report Card Grades." In *Student-Centered Classroom Assessment*, pp. 363–396. New York: Macmillan.

Stiggins, R. J., and D. L. Duke. (1991). "District Grading Policies and Their Potential Impact on At-risk Students." Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Stiggins, R. J., D. A. Frisbie, and P. A. Griswold. (1989). "Inside High School Grading Practices: Building a Research Agenda."*Educational Measurement: Issues and Practice* 8, 2: 5–14.

Sweedler-Brown, C. O. (1992). "The Effect of Training on the Appearance Bias of Holistic Essay Graders." *Journal of Research and Development in Education* 26, 1: 24–29.

Wright, R. G. (1994). "Success for All: The Median Is the Key." *Phi Delta Kappan* 75, 9: 723–725.