

University of Kentucky
UKnowledge

Educational, School, and Counseling Psychology Faculty Publications

Educational, School, and Counseling Psychology

11-2011

Five Obstacles to Grading Reform

Thomas R. Guskey
University of Kentucky, GUSKEY@UKY.EDU

Follow this and additional works at: https://uknowledge.uky.edu/edp_facpub

 Part of the Educational Assessment, Evaluation, and Research Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

Guskey, Thomas R., "Five Obstacles to Grading Reform" (2011). *Educational, School, and Counseling Psychology Faculty Publications*. 6.
https://uknowledge.uky.edu/edp_facpub/6

This Article is brought to you for free and open access by the Educational, School, and Counseling Psychology at UKnowledge. It has been accepted for inclusion in Educational, School, and Counseling Psychology Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Five Obstacles to Grading Reform

Notes/Citation Information

Published in *Educational Leadership*, v. 69, issue 3, p. 16-21.

Copyright © 2011 Thomas R. Guskey

The copyright holder has granted the permission for posting the article here.

This article is available at UKnowledge: https://uknowledge.uky.edu/edp_facpub/6

Five Obstacles to



Grading Reform

Education leaders must recognize obstacles to grading reform that are rooted in tradition—and then meet them head on.

Thomas R. Guskey

Education improvement efforts over the past two decades have focused primarily on articulating standards for student learning, refining the way we assess students' proficiency on those standards, and tying results to accountability. The one element still unaligned with these reforms is grading and reporting. Student report cards today look much like they looked a century ago, listing a single grade for each subject area or course.

Educators seeking to reform grading must combat five long-held traditions that stand as formidable obstacles to change. Although these traditions stem largely from misunderstandings about the goals of education and the purposes of grading, they remain ingrained in the social fabric of our society.

Obstacle 1:

Grades should provide the basis for differentiating students.

This is one of our oldest traditions in grading. It comes from the belief that grades should serve to differentiate students on the basis of demonstrated talent. Students who show superior talent receive high grades, whereas those who display lesser talent receive lower grades.

Although seemingly innocent, the implications of this belief are significant and troubling. Those who enter the profession of education must answer one basic, philosophical question: Is my purpose to *select* talent or *develop* it? The answer must be one or the other because there's no in-between.

If your purpose as an educator is to *select* talent, then you must work to maximize the differences among students. In other words, on any measure of learning, you must try to achieve the greatest possible variation in students' scores. If students' scores on any measure of learning are clustered closely together, discriminating among them becomes difficult, perhaps even impossible. Unfortunately for students, the best means of maximizing differences in learning is poor teaching. Nothing does it better.

Assessments also play a role. Assessments used for selection purposes, such as college entrance examinations like the ACT and SAT, are designed to be instructionally insensitive (Popham, 2007). That is, if a particular concept is taught well and, as a result, most students answer an assessment item related to that concept correctly, it no longer discriminates among students and is therefore eliminated from the assessment. These types of assessments maximize differences among students, thus facilitating the selection process.

If, on the other hand, your purpose as an educator is to *develop* talent, then you go about your work differently. First, you clarify what you want students to learn and be able to do. Then you do everything possible to ensure that *all* students learn those things well. If you succeed, there should be little or no variation in measures of student learning. All students are likely to attain high scores on measures of achievement, and all might receive high grades. If your purpose is to develop talent, this is what you strive to accomplish.

© BENJAMIN RONDEL/CORBIS

Obstacle 2:

Grade distributions should resemble a normal bell-shaped curve.

The reasoning behind this belief goes as follows: If scores on intelligence tests tend to resemble a normal bell-shaped curve—and intelligence is clearly related to achievement—then grade distributions should be similar.

A true understanding of normal curve distributions, however, shows the error in this kind of reasoning. The normal bell-shaped curve describes the distribution of randomly occurring events *when nothing intervenes*. If we conducted an experiment on crop yield in agriculture, for example, we would expect the results to resemble a normal curve. A few fertile fields would produce a high yield; a few infertile fields would produce a low yield; and most would produce an average yield, clustering around the center of the distribution.

But if we intervene in that process—say we add a fertilizer—we would hope to attain a very different distribution of results. Specifically, we would hope to have all fields, or nearly all, produce a high yield. The ideal result would be for all fields to move to the high end of the distribution. In fact, if the distribution of crop yield after our intervention still resembled a normal bell-shaped curve, that would show that our intervention had failed because it made no difference.

Teaching is a similar intervention. It's a purposeful and intentional act. We engage in teaching to attain a specific result—that is, to have all students, or nearly all, learn well the things we set out to teach. And just like adding a fertilizer, if the distribution of student learning after teaching resembles a normal bell-shaped curve, that, too, shows the degree to which our intervention failed. It made no difference.

Further, research has shown that the seemingly direct relationship between aptitude or intelligence and school achievement depends on instructional conditions, *not* a normal distribution curve (Hanushek, 2004; Hershberg, 2005). When the instructional quality is high and well matched to students' learning needs, the magnitude of the relationship between aptitude/intelligence and school achievement diminishes drastically and approaches zero (Bloom, 1976; Bloom, Madaus, & Hastings, 1981).

If someone proposed combining measures of height, weight, diet, and exercise into a single number to represent a person's physical condition, we would consider it laughable.

Obstacle 3:

Grades should be based on students' standing among classmates.

Most parents grew up in classrooms where their performance was judged against that of their peers. A grade of *C* didn't mean you had reached Step 3 in a five-step process to mastery or proficiency. It meant "average" or "in the middle of the class." Similarly, a high grade did not necessarily represent excellent learning. It simply meant that you did better than most of your classmates. Because most parents experienced such norm-based grading procedures as children, they see little reason to change them.

But there's a problem with this approach: Grades based on students' standing among classmates tell us nothing about how well students have learned. In such a system, all students might have performed miserably, but some simply performed less miserably than others.

In addition, basing grades on students' standing among classmates makes learning highly competitive. Students must compete with one another for the few scarce rewards (high grades) to be awarded by teachers. Doing well does not mean learning excellently; it means outdoing your classmates. Such competition damages relationships in school (Krumboltz & Yeh, 1996). Students are discouraged from cooperating or helping one another because doing so might hurt the helper's chance at success. Similarly, teachers may refrain from helping individual students because some students might construe this as showing favoritism and biasing the competition (Gray, 1993).

Grades must always be based on clearly specified learning criteria. Those criteria should be rigorous, challenging, and transparent. Curriculum leaders who are working to align instructional programs with the newly developed common core state standards move us in that direction. Grades based on specific learning criteria have direct meaning; they communicate what they were intended to communicate.

Obstacle 4:

Poor grades prompt students to try harder.

Although educators would prefer that motivation to learn be entirely intrinsic, evidence indicates that grades and other reporting methods affect student motivation and the effort students put forth (Cameron & Pierce, 1996). Studies show that most students view high grades as positive recognition of their success, and some work hard to avoid the consequences of low grades (Haladyna, 1999).

At the same time, no research

supports the idea that low grades prompt students to try harder. More often, low grades prompt students to withdraw from learning. To protect their self-images, many students regard the low grade as irrelevant or meaningless. Others may blame themselves for the low grade but feel helpless to improve (Selby & Murphy, 1992).

Recognizing the effects on students of low grades, some schools have initiated policies that eliminate the use of failing grades altogether. Instead of assigning a low or failing grade, teachers assign an *I*, or incomplete, with immediate consequences. Students who receive an *I* may be required to attend a special study session *that day* to bring their performance up to an acceptable level—and no excuses are accepted. Some schools hold this session after regular school hours whereas others conduct it during lunchtime.

Such a policy typically requires additional funding for the necessary support mechanisms, of course. But in the long run, the investment can save money. Because this regular and ongoing support helps students remedy their learning difficulties before they become major problems, schools tend to spend less time and fewer resources in major remediation efforts later on (see Roderick & Camburn, 1999).

Obstacle 5:

Students should receive one grade for each subject or course.

If someone proposed combining measures of height, weight, diet, and exercise into a single number or mark to represent a person's physical condition, we would consider it laughable. How could the combination of such diverse measures yield anything meaningful? Yet every day, teachers combine aspects of students' achievement, attitude, responsibility, effort, and behavior into a single grade that's recorded on a report card—and no one questions it.

In determining students' grades, teachers typically merge scores from major exams, compositions, quizzes, projects, and reports, along with evidence from homework, punctuality in turning in assignments, class participation, work habits, and effort. Computerized grading programs help teachers apply different weights to each of these categories (Guskey, 2002a) that then are combined in idiosyncratic ways (see McMillan, 2001; McMillan, Myran, & Workman, 2002). The result is a "hodgepodge grade" that is just as confounded and impossible to interpret as a "physical condition" grade that combined height, weight, diet, and exercise would be (Brookhart & Nitko, 2008; Cross & Frary, 1996).

Recognizing that merging these diverse sources of evidence distorts the meaning of any grade, educators in many parts of the world today assign multiple grades. This idea provides the foundation for standards-based approaches to grading. In particular, educators distinguish product, process, and progress learning criteria (Guskey & Bailey, 2010).

Product criteria are favored by educators who believe that the primary purpose of grading is to communicate summative evaluations of students' achievement and performance (O'Connor, 2002). In other words, they focus on *what* students know and are able to do at a particular point in time. Teachers who use product criteria typically base grades exclusively on final examination scores; final products (reports, projects, or exhibits); overall assessments; and other culminating demonstrations of learning.

Process criteria are emphasized by educators who believe that product criteria do not provide a complete picture of student learning. From their perspective, grades should reflect not only the final results, but also *how* students got there. Teachers who consider responsibility, effort, or work habits when assigning grades use process criteria.

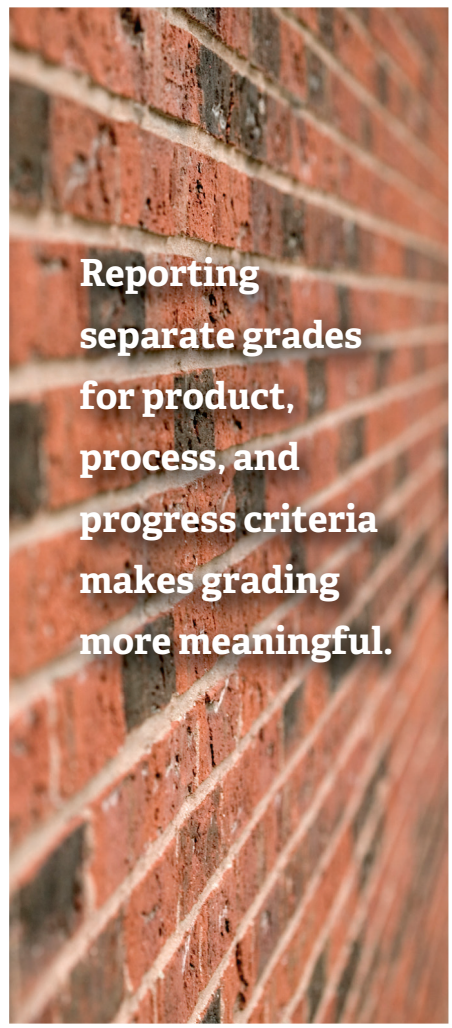
Grades based on students' standing among classmates tell us nothing about how well students learned.

So do teachers who count classroom quizzes, formative assessments, homework, punctuality of assignments, class participation, or attendance.

Progress criteria are used by educators who believe that the most important aspect of grading is how much students gain from their learning experiences. Other names for progress criteria include *learning gain*, *improvement scoring*, *value-added learning*, and *educational growth*. Teachers who use progress criteria look at how much improvement students have made over a particular period of time, rather than just where they are at a given moment. As a result, scoring criteria may be highly individualized among students. Grades might be based, for example, on the number of skills or standards in a learning continuum that students mastered and on the adequacy of that level of progress for each student. Most of the research evidence on progress criteria comes from studies of individualized instruction (Esty & Teppo, 1992) and special education programs (Gersten, Vaughn, & Brengelman, 1996; Jung & Guskey, 2010).

After establishing explicit indicators of product, process, and progress learning criteria, teachers in countries that differentiate among these indicators assign separate grades to each indicator. In this way, they keep grades for responsibility, learning skills, effort, work habits, or learning progress distinct from assessments of achievement and performance (Guskey, 2002b; Stiggins, 2008). The intent is to provide a more accurate and comprehensive picture of what students accomplish in school.

Although schools in the United States are just beginning to catch on to the idea of separate grades for product, process, and progress criteria, many Canadian educators have used the practice for years (Bailey & McTighe, 1996). Each marking period, teachers in these schools assign an achievement grade on the basis of the student's performance on projects, assessments, and other demonstrations of learning. Often expressed as a letter grade or percentage (*A* = advanced, *B* = proficient, *C* = basic, *D* = needs improvement, *F* = unsatisfactory), this achievement grade represents the teacher's judgment of the student's level of performance relative to explicit learning goals established for the subject area or course. Computations of grade-point averages and class ranks are based solely on these achievement or "product" grades.



In addition, teachers assign separate grades for homework, class participation, punctuality of assignments, effort, learning progress, and the like. Because these factors usually relate to specific student behaviors, most teachers record numerical marks for each (*4* = consistently; *3* = usually; *2* = sometimes; and *1* = rarely). To clarify a mark's meaning, teachers often identify specific behavioral indicators. For example, these might be the indicators for a homework mark:

- *4* = All homework assignments are completed and turned in on time.
- *3* = There are one or two missing or incomplete homework assignments.
- *2* = There are three to five missing or incomplete homework assignments.
- *1* = There are numerous missing or incomplete homework assignments.

Teachers sometimes think that reporting multiple grades will increase their grading workload. But those who use the procedure claim that it actually makes grading easier and less work (Guskey, Swan, & Jung, 2011a). Teachers gather the same evidence on student learning that they did before, but they no longer worry about how to weigh or combine that evidence in calculating an overall grade. As a result, they avoid irresolvable arguments about the appropriateness or fairness of various weighting strategies.

Reporting separate grades for product, process, and progress criteria also makes grading more meaningful. Grades for academic achievement reflect precisely that—academic achievement—and not some confusing amalgamation that's impossible to interpret and that rarely presents a true picture of students' proficiency (Guskey, 2002a). Teachers also indicate that students take homework more seriously when it's reported separately. Parents favor the practice because it provides a more comprehensive profile of their child's performance in school (Guskey, Swan, & Jung, 2011b).

The key to success in reporting multiple grades, however, rests in the clear specification of indicators related to product, process, and progress criteria.

Teachers must be able to describe how they plan to evaluate students' achievement, attitude, effort, behavior, and progress. Then they must clearly communicate these criteria to students, parents, and others.

No More "We've Always Done It That Way"

Challenging these traditions will not be easy. They've been a part of our education experiences for so long that they usually go unquestioned, despite the fact that they are ineffective and potentially harmful to students.

Education leaders who challenge these traditions must be armed with thoughtful, research-based alternatives. You can't go forward with only passionately argued opinions. To succeed in tearing down old traditions, you must have new traditions to take their place.

No research supports the idea that low grades prompt students to try harder. More often, low grades prompt students to withdraw from learning.

This means that education leaders must be familiar with the research on grading and what works best for students so they can propose more meaningful policies and practices that support learning and enhance students' perceptions of themselves as learners. Leaders who have the courage to challenge the traditional approach and the conviction to press for thoughtful, positive reforms are likely to see remarkable results. EL

References

Bailey, J. M., & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. In T. R. Guskey (Ed.), *Communicating student learning: 1996 Yearbook of the Association for Supervision and Curriculum Development* (pp. 119–140). Alexandria, VA: ASCD.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. New York: McGraw-Hill.

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson.

Cameron, J., & Pierce, W. D. (1996). The debate about rewards and intrinsic motivation: Protests and accusations do not alter the results. *Review of Educational Research*, *66*(1), 39–51.

Cross, L. H., & Frary, R. B. (1996, April). *Hodgepodge grading: Endorsed by students and teachers alike*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Esty, W. W., & Teppo, A. R. (1992). Grade assignment based on progressive improvement. *Mathematics Teacher*, *85*(8), 616–618.

Gersten, R., Vaughn, S., & Brengelman, S. U. (1996). Grading and academic feedback for special education students and students with learning difficulties. In T. R. Guskey (Ed.), *Communicating student learning: 1996 yearbook of the Association for Supervision and Curriculum Development* (pp. 47–57). Alexandria, VA: ASCD.

Gray, K. (1993). Why we will lose: Taylorism in America's high schools. *Phi Delta Kappan*, *74*(5), 370–374.

Guskey, T. R. (2002a). Computerized gradebooks and the myth of objectivity. *Phi Delta Kappan*, *83*(10), 775–780.

Guskey, T. R. (2002b). *How's my kid doing? A parents' guide to grades, marks, and report cards*. San Francisco: Jossey Bass.

Guskey, T. R., & Bailey, J. M. (2010). *Developing standards-based report cards*. Thousand Oaks, CA: Corwin.

Guskey, T. R., Swan, G. M., & Jung, L. A. (2011a, April). *Parents' and teachers' perceptions of standards-based and traditional report cards*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Guskey, T. R., Swan, G. M., & Jung, L. A. (2011b). Grades that mean something: Kentucky develops standards-based report cards. *Phi Delta Kappan*, *93*(2).

Haladyna, T. M. (1999). *A complete guide to student grading*. Boston: Allyn and Bacon.

Hanushek, E. A. (2004). *Some simple analytics of school quality* (Working paper 10229). Cambridge, MA: National Bureau of Economic Quality.

Hershberg, T. (2005). Value-added assessment and systemic reform: A response to the challenge of human capital development. *Phi Delta Kappan*, *87*(4), 276–283.

Jung, L. A., & Guskey, T. R. (2010). Grading exceptional learners. *Educational Leadership, 67*(5), 31–35.

Krumboltz, J. D., & Yeh, C. J. (1996). Competitive grading sabotages good teaching. *Phi Delta Kappan*, *78*(4), 324–326.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20–32.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Educational Research*, *95*(4), 203–213.

O'Connor, K. (2002). *How to grade for learning: Linking grades to standards* (2nd ed.). Arlington Heights, IL: SkyLight.

Popham, J. W. (2007). Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan*, *89*(2), 146–150.

Roderick, M., & Camburn, E. (1999). Risk and recovery from course failure in the early years of high school. *American Educational Research Journal*, *36*(2), 303–343.

Selby, D., & Murphy, S. (1992). Graded or degraded: Perceptions of letter-grading for mainstreamed learning-disabled students. *British Columbia Journal of Special Education*, *16*(1), 92–104.

Stiggins, R. J. (2008). Report cards: Assessments *for* learning. In *Student-involved assessment for learning* (5th ed., pp. 267–310). Upper Saddle River, NJ: Merrill/Prentice.

Copyright © 2011 Thomas R. Guskey

Thomas R. Guskey is professor in the Department of Educational, School, and Counseling Psychology, College of Education, University of Kentucky, Lexington; 859-257-5748; guskey@uky.edu.