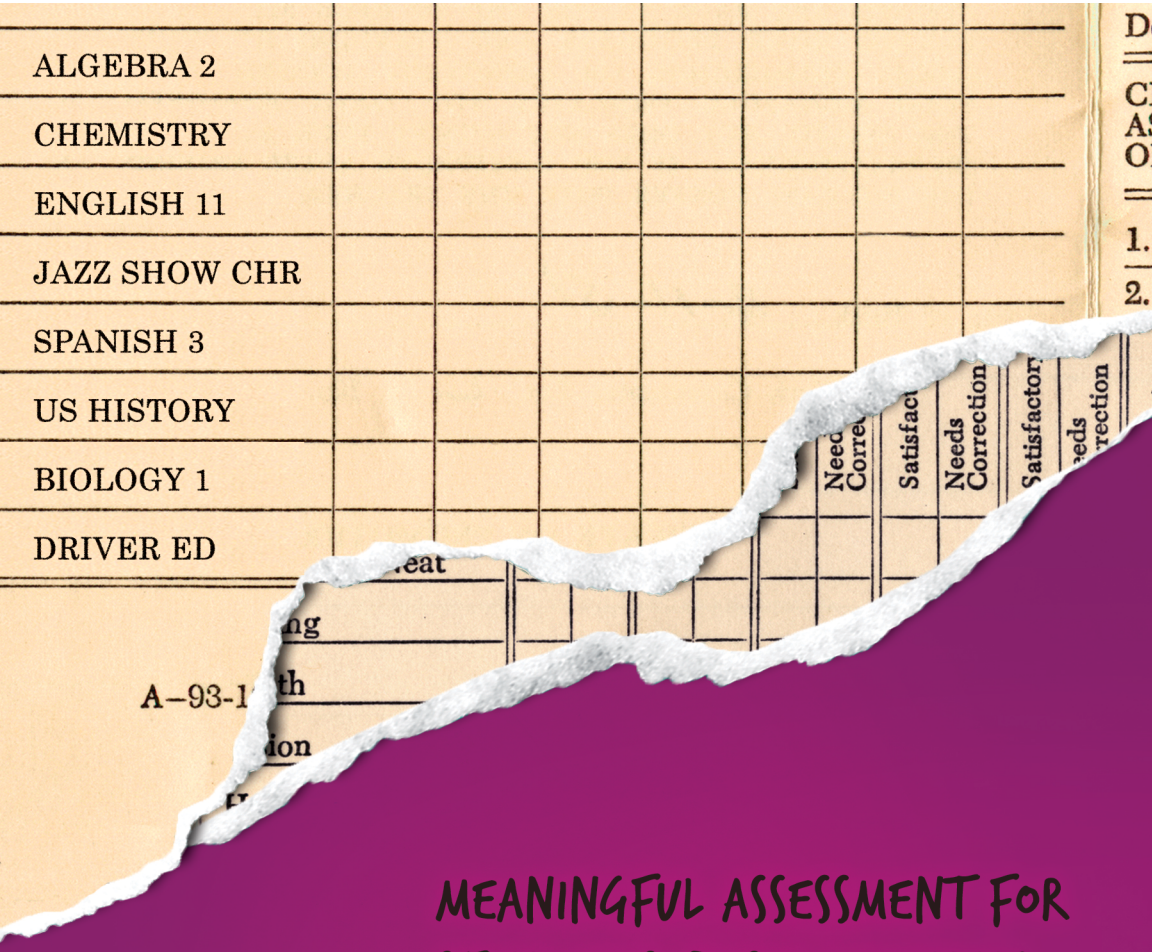

MEANINGFUL ASSESSMENT FOR STANDARDS-BASED LEARNING

Rethinking Grading

Cathy Vatterott

1703 N. Beauregard St. • Alexandria, VA 22311-1714 USA
Phone: 800-933-2723 or 703-578-9600 • Fax: 703-575-5400
Website: www.ascd.org • E-mail: member@ascd.org
Author guidelines: www.ascd.org/write

Judy Seltz, *Executive Director,* Stefani Roth, *Publisher;* Genny Ostertag, *Director, Content Acquisitions;* Julie Houtz, *Director, Book Editing & Production;* Darcie Russell, *Senior Associate Editor;* Lindsey Smith, *Senior Graphic Designer;* Mike Kalyan, *Manager, Production Services;* Valerie Younkin, *Production Designer;* Kyle Steichen, *Senior Production Specialist*

PAPERBACK ISBN: 978-1-4166-2049-5 ASCD product #115001

PDF E-BOOK ISBN: 978-1-4166-2051-8; see Books in Print for other formats.

Quantity discounts: 10–49, 10%; 50+, 15%; 1,000+, special discounts (e-mail programteam@ascd.org or call 800-933-2723, ext. 5773, or 703-575-5773). For desk copies, go to www.ascd.org/deskcopy.

ASCD Member Book No. FY15-8A (July 2015 PSI+). ASCD Member Books mail to Premium (P), Select (S), and Institutional Plus (I+) members on this schedule: Jan, PSI+; Feb, P; Apr, PSI+; May, P; Jul, PSI+; Aug, P; Sep, PSI+; Nov, PSI+; Dec, P. For current details on membership, see www.ascd.org/membership.

Library of Congress Cataloging-in-Publication Data
TK

23 22 21 20 19 18 17 16 15 14 1 2 3 4 5 6 7 8 9 10 11 12

Rethinking Grading

MEANINGFUL ASSESSMENT FOR STANDARDS-BASED LEARNING

Acknowledgments

My thanks go first to ASCD for an educational vision that has guided my career and for the forum it has provided for my ideas. I am honored to be an ASCD author. I also wish to thank Genny Ostertag, who served as my editor, coach, and taskmaster. She deftly navigated me through a series of twists, turns, delays, and rewrites to shape my ideas into a coherent manuscript. Her unrelenting encouragement and patience were invaluable. Thanks also to Darcie Russell, for smoothing the rough edges from my sometimes irreverent style and for her meticulous attention to detail.

I am beholden to all the authors who dared to question our most entrenched grading practices and whose research formed the foundation for the work of so many others. I have been particularly influenced by Alfie Kohn, Robert Lynn Canady, Tom Guskey, Ken O'Connor, Doug Reeves, and Tom Schimmer. Their writings, presentations, and personal conversations intrigued me, compelled me to reflect, and brought clarity to the big picture of grading.

Next, my thanks go out to those educators who befriended me during my research. I appreciate the thoughtfulness of my "grading buddies" Jeff Harding, Rob McEntarffer, Hugh O'Donnell, Matt Townsley, and Nathan Wear, who regularly took the time to chat

with me and to share web links and articles they thought would help me. They remain great colleagues who have made the journey fun and exciting.

Finally, I am especially grateful to all the K–12 teachers and administrators who willingly gave me access to their expertise, their classrooms, and their grade books. I remain in awe of their energy, caring, and commitment to the students. They unselfishly gave of their time to teach me what standards-based grading looked like, and brought the concepts to life by sharing their practices. Their passion for standards-based grading was infectious and my visits with them never failed to energize me. They were my inspiration—without them, this book would not be possible.

Introduction

Grading. It's the hardest decision I must make in teaching. Every semester I agonize about what is right. Is the grade truly indicative of the student's performance? Do I have the weighting right? Have I given students enough opportunities to improve? All this happens within the confines of the amount of time and energy I have available to determine grades.

I continue to evolve in my practice. Last semester I created structured formative activities and rubrics. I found that those strategies helped more students "hit the mark" the first time. I had fewer rewrites with fewer deficiencies.

Writing this book has been an awakening. Each semester I am more thoughtful, more analytical, and more reflective about my own grading practices. At the same time, I am troubled by the mindset that my college freshmen and sophomores bring to my courses. (The juniors and seniors are a bit better.) For many, their K–12 experience has left them woefully unprepared to handle college-level work. They often don't fully comprehend how to analyze and synthesize. They seem to be stuck in the mode of "just tell me what you want." Many of them are terrible writers, unable to express their thoughts clearly and intelligently. I see firsthand the damage we have done, and how we have handicapped them

for college by giving grades that don't reflect learning. I am not surprised by how many drop out.

We have the opportunity to change that. We have the opportunity to send students forward to college with the conceptual knowledge and learning strategies that are so critical to college success. It starts with empowering them to take charge of their own learning and by eliminating some of the obstacles of traditional grading. Standards-based grading has the potential to restore integrity to the grading process. It can and will change our students' futures.

1

The Culture of Grading

People are doing a lot of rethinking about education these days. The pundits agree that something is wrong with K–12 education and everyone has a solution: a longer school day, a longer school year, more testing, less recess. Columnists, talk show hosts, and politicians on both sides lament that we've lost our edge. Competition is global and according to the tests we are not keeping up.

Why does global competitiveness matter? The intersection of globalization and technology has created an international competition for jobs and even college admissions. We can now easily compete, connect, and collaborate with people around the world (Friedman & Mandelbaum, 2012). "In today's interconnected world, our students are not competing with students from the state or city next door, but with students from Singapore, Shanghai, and Stockholm" (Stewart, 2012, p. 3). To be average is no longer good enough.

How are we doing? Some say there is a crisis in U.S. education. Others say the crisis is overblown. But there are some indisputable facts. On international tests, our students are performing poorly compared with students from other countries. Three international tests compare math, science, and reading

performance—Trends in International Mathematics and Science Survey (TIMMS), Progress in International Reading Literacy Study (PIRLS), and the Programme for International Student Assessment (PISA). In 2011, U.S. 8th graders came in 7th place in math and 9th place in science on the TIMMS. In that same year, U.S. 4th graders ranked 6th out of 53 countries in reading on the PIRLS. PISA is the most widely used international test, measuring performance in 65 countries. It is also the most challenging in that its goal is to measure not merely content knowledge, but the ability of students to apply knowledge to solve real-life problems. On the PISA in 2012, U.S. students scored well below other developed countries—23rd in science, 30th in math, and 20th in reading (U.S. Department of Education, 2012a; 2012b; 2012c).

Even if we discount standardized test scores as an indicator of how our students are doing, we know this: In the United States, we not only have a *skills gap* (jobs that can't be filled due to a lack of skilled labor) but also a *learning gap* (an unacceptable high school and college dropout rate as well as college students who need remediation). The skills gap is evident in the lack of workers with specific skills needed for some of today's jobs, jobs that did not exist only a few years ago. In early 2012, in spite of the recession, there were more than three million jobs vacant in the United States due to a lack of math, reading comprehension, or technical skills required by companies (Friedman & Mandelbaum, 2012). The learning gap is twofold. First, 25 percent of our students will not graduate high school. The high school graduation rate in the United States of 74.7 percent ranks 12th among 28 developed countries (*Education Week*, 2013; Stewart, 2012). A high school diploma qualifies graduates for only a few low-wage jobs; adults without a high school diploma face dismal job prospects. Second, if U.S. students *do* make it to college, one-third of them must take at least one remedial course in reading, writing, or math. Only 54 percent of those entering college in the U.S. will complete a degree, ranking near the bottom when compared to other country's rates

of college completion. Slots at elite universities are increasingly filled by better prepared students from other countries.

People bemoan the sad state of U.S. competitiveness and insist that education is both the cause and the fix of our woes. Yet no one seems to have a definitive answer about what in education needs to be fixed. No one has the answer because there isn't just one answer. It's a series of related problems that overlap between curriculum, instruction, and assessment.

Although K–12 educational reform is not the cure-all for the ills of the United States, the reform of one educational practice—grading—has the potential to drive related changes in other practices. The culture of grading and all the baggage it encompasses has perpetuated a system that obstructs many other educational reforms. What is the relationship between grading reform and overall educational reform? Are grades a *reflection* of a dysfunctional system or a *driver* of the system? Hard to say. Grades are supposed to reflect what students know and how well the teacher has taught. But they often don't.

We now know that something is wrong with grades. Every day we see the mismatch—on one hand, the stellar performance on standardized tests from *B* and *C* students (thus labeled "underachievers"), and on the other, poor performance on standardized tests from straight *A* students. We know that many students leave high school with high grade point averages yet struggle academically in college.

Let us reflect for a moment on the "what ifs"—that perhaps the answer lies in the reform of a traditional educational practice that has not changed in decades. What if grading practices were a piece of a bigger picture? What if by changing the way we use grades we could ignite authentic high-level learning? What if student empowerment could make learning more dynamic and change the outcome? What if our beliefs about grading were misguided?

If we dare to question our beliefs about grading, more "what if" questions emerge.

- What if an *A* student was a compliant one rather than a learned one?
- What if the premise that high grades were a predictor of success in life was faulty?
- What if grades, as the marker of success in school, were a flawed, or worse yet, meaningless tool?
- What if parents, by directing their children to focus on grades, inadvertently created an addiction to form over substance?

The challenge of reforming grading practices is a difficult one. The "what ifs" reveal a practice that is deeply ingrained not only in education but in our culture. Grading is a language, a schema—we grade presidents and we grade meat. For grading reform to happen, we must acknowledge and accept how our beliefs have influenced grading practices.

A Brief History of Education

How did we get here? Three historical forces converged to create and perpetuate traditional grading practices that are common today—the roots of education in moral development, the use of education to sort and rank students, and the prevalence of behaviorism in school practices.

Teacher as Moral Educator

In a young and often chaotic colonial America, moral stability was necessary for the survival of society. The original establishment of schools was primarily for the purpose of moral education, and schools were viewed as an important social agency to promote virtue, character, and good habits. From the earliest days of our country, the goal of mass literacy was driven by the need to

read the Bible and thus, save one's soul. Contrary to today's practice of secular education, schools were the servant of religion, and moral education in the schools was a logical outgrowth of religion. "In the eyes of Puritans religious and moral education were inextricably intertwined" (McClellan, 1999, p. 2). Learning was valued not as an end in itself but as "an instrument for clarifying the ways of God to man and thus rendering certain the conditions of eternal salvation" (Thayer, 1965, p. 12). Teachers worked hard to promote in students the virtues of self-restraint, industry, honesty, punctuality, and orderliness. Discipline in school was viewed as a way to model full obedience to God.

A basic fear of the fragility of human virtue pervaded our society—that without constraints and vigilance our youth would fall prey to unsavory temptations. This fear was grounded in the 17th century conception of original sin, that man was predisposed to choose evil over good (Thayer, 1965). The fears of our founding fathers were not much different from the general concerns for our youth today and the roots of moral education are evident in today's common educational practices. We reward the modern version of virtue and punish the lack of it. We reward responsibility, effort, hard work, neatness, and homework completion. We penalize tardiness, sloppiness, late work, and cheating. For this noble goal of instilling morality in students, grades have been a most convenient tool. Unfortunately, this use of grades has led to a school culture that often places more value on compliance and working than learning.

Schools as a Mechanism for Sorting and Ranking

Early in the 20th century, compulsory attendance laws changed the practice of K–12 education in the United States. Elementary schools grew in popularity and large numbers of students started attending high school. From 1870 to 1910 the number of high schools in the United States grew from 500 to 10,000, and the total number of students in public elementary and high

schools grew from 6,871,000 to 17,813,000 (Kirschenbaum, Simon, & Napier, 1971). While elementary schools continued to report student learning with narratives, the sheer number of students at the secondary level made such descriptive reports burdensome. Secondary schools, eager for a more efficient alternative, began examining techniques used in colleges.

In the late 1700s, Yale was probably the first college to rank student performance into four categories, a practice that evolved into the use of a four-point scale (a precursor of the four-point grade point average). In 1877, Harvard began classifying students using percentages, which was later replaced by classifying students into five groups, the lowest of which failed the class (Durm, 1993). In 1897, Mount Holyoke College adopted a system that combined descriptive adjectives with percentages and letters:

A = Excellent, equivalent to 95–100 percent
B = Good, equivalent to 85–94 percent
C = Fair, equivalent to 76–84 percent
D = Passed (barely), equivalent to 75 percent
E = Failed (below 75 percent) (Durm, 1993, p. 3)

This system of grading, with variations from school to school, evolved to become the standard for sorting and ranking college students and was soon adopted by secondary schools. Letter grades were an easy, efficient method not only for telling students how they were doing, but also for ability grouping students for instruction. As the number of high school students applying to college increased, colleges starting using high school grades to screen applicants.

In 1912, some powerful research emerged about the lack of consistency in percentage grades. When English exams from two students were scored by 142 different teachers, the scores on one exam ranged from 64 to 98 percent and scores on the other exam ranged from 50 to 97 percent. The same experiment with geometry

papers showed even more discrepancy, with the grades ranging from 28 to 95 percent (Starch & Elliott, 1912; Starch & Elliott, 1913). This research was viewed as so damaging to the practice of using percentages that educators began moving away from the 100-point scale to the five categories of *A*, *B*, *C*, *D*, *F*. Fewer categories seemed more "fair."

Around this time, a new method became popular—grading on the curve. The bell curve, technically called the *normal distribution*, can actually be traced back to the work of statisticians and mathematicians as early as the 18th century. It became popular in the 20th century after it was shown that many physical and psychological phenomena (such as height) presented as a normal distribution. The use of the bell curve in education became popular when IQ scores of a random group of children were shown to fall into a bell-shaped curve (Curreton, 1971; Jensen, 1980). Grading on the curve was believed to be appropriate because, at that time, the distribution of students' intelligence test scores approximated a normal probability curve. Since innate intelligence and school achievement were thought to be directly related, such a procedure seemed both fair and equitable (Guskey, 1996). The assumption that individual aptitude was fixed and that aptitude varied among students led to the logical conclusion that achievement should also present as a bell curve. Regardless of its validity, the bell curve became popular as a way to produce a "fair" distribution of grades. For the purpose of sorting and ranking students, the bell curve was ideal.

Upon closer investigation, however, the logic was shown to be faulty. The normal curve was not considered statistically valid unless the group was large, random, and untreated (Kelly, 2009). "The normal bell-shaped curve describes the distribution of randomly occurring events *when nothing intervenes*" (Guskey, 2011, p. 18). But teachers intervene—they teach with the goal of having all students learn. "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" (Guskey, 2011, p. 18). More recent research has also shown that the relationship between aptitude/intelligence and school achievement is dependent upon the appropriateness of instructional conditions (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" (Guskey, 2011, p. 18). In spite of this fact, the bell curve is still mistakenly revered today as evidence of rigor.

Behaviorism as a Tool for Compliance

If morality and sorting and ranking were what we wanted, behaviorism was the way to reach the goal.

> You may find the roots of behaviorism and its counterpart in education, behavior modification, to have germinated from some unusual sources—a salivating dog, a ringing bell, and, more recently, some chocolate covered candies. (Freiberg, 1999, p. 5)

Behaviorism, a major contribution to the field of psychology, dates back to the late 17th century. At that time, Edward Thorndike theorized about the Law of Effect, that behavior leading to a positive consequence will be repeated (Kohn, 1999). Skinner's theory of operant conditioning showed that human behavior could be shaped through positive reinforcement. For example, the practice of placing chocolate candies on students' desks was used to reinforce good behavior (Freiberg, 1999).

Pavlov learned that he could get a dog to salivate simply by associating the sound of a ringing bell with food. Pavlov won the Nobel Prize in 1904 for his study of digestion, but the research was applied to a new psychological theory for humans (Freiberg, 1999). As behaviorism grew in popularity, it led to a new approach to classroom management. Behavior modification emerged in the

1960s and rapidly became the dominant philosophy for classroom management. One of the more popular programs was Lee Canter's Assertive Discipline.

Canter's Assertive Discipline model was a structure of rewards and consequences that were used to control student behavior. The most familiar consequence for bad behavior was that of writing the student's name on the board, followed by check marks behind the name for each additional infraction. The name on the board was meant as a warning, the check mark as a consequence. "If you do break this rule again, or any other rule during the day, I'll put a check next to your name . . . this means that you have chosen to sit for five minutes in the time-out area" (Canter & Canter, 1992, p. 103). A popular reward for good behavior was Marbles in a Jar.

> When the class is doing what you want, you take a marble and drop it in a jar. The sound of the marble dropping into the jar immediately lets the students know they are doing what you want and that you recognize their efforts. Each marble can be worth, for example, 30 seconds to one minute of free choice at the end of the day. (Canter, 1976, p. 141)

The widespread use of behavior modification for classroom management generalized to other school practices such as detentions for misbehavior, awards for perfect attendance, and even the use of bells (particularly ironic in light of Pavlov). Behavior management became the dominant paradigm in schools for controlling the *behavior* of learning (or so we thought) as well as controlling classroom behavior.

Today the idea that behavior can be controlled by rewards and punishment is so embedded in the day-to-day practices of schools, one rarely even notices it (Kohn, 1999). Given the fact that "behaviorism permeates virtually every aspect of American education" (Kohn, 1999, p. 143), it is no surprise that grades have become a major tool for rewarding good. So although grades originally

evolved for the purpose of sorting and ranking, they turned out to be quite handy for rewarding virtue and punishing vice. They were an all-purpose tool—not only for rewarding achievement but behaviors such as compliance and responsibility.

Beliefs About Grading

From the historical forces of morality, sorting, and behaviorism, a culture of grading evolved that was a mix of moralistic views of human nature, the puritan work ethic, and the use of reward and punishment to shape behavior. The culture is evident in a set of beliefs about grading, the quasi-superstitions that drive educational practice. Those beliefs developed from the most honorable motives. As educators we have been concerned not only about intellectual growth, but also moral development and the preparation of children for adulthood. We've used grades for more than academics because we believe our job is more than academics—our goals have always included shaping children into better people. But our well-meaning beliefs and their unintended consequences deserve closer examination.

Belief #1: Good Teachers Give Bad Grades

As we accepted that the role of the school was to sort and rank, we came to believe that in a rigorous educational system, success was scarce. Scarcity of high grades equaled rigor and only a few should be "winners." (A student once said to her professor, "Well if everyone got an *A*, then it doesn't mean anything.") From a practical standpoint, we also realized if there were too many high grades, sorting and ranking students would be difficult.

The bell curve became the foundation for academic competition—it became synonymous with academic rigor. We accepted the bell curve as an indication of some natural order of things. We were confident our belief in it was valid, because, after all, the bell curve was science. If the purpose of grading was to rank and

discriminate poor learners from good learners, then the closer the grades approximated a bell-shaped curve the better. If we set the standard for excellence so high that only a few students excelled and many failed, then we must have very high standards. Then, of course, the inverse must be true—lots of high grades must indicate low standards and a watered-down curriculum. (As a first year teacher, I was called to the principal's office to be chastised for giving too many *A*s, until I explained that I taught for mastery, with formative feedback and retakes).

As teachers, we bought into the idea that a bell curve indicated rigor and misinterpreted it to be a *rule to follow* (any idea can be bastardized!). We came to believe that if success were scarce and grades fell into a bell curve, then we were tough teachers. The grades became our yardstick—a measure of how tough we were and a measure of how tough the content was. This belief became so commonplace that nearly all college students have had the experience on the first day of class when the professor proudly announces "half of you will fail this class." When enough students did not fail, the teacher or the curriculum was labeled "too easy" (Vatterott, 2007). We even have a name for it—grade inflation.

In 1894**,** only a decade after Harvard adopted letter grades, a group of professors began complaining that "grades *A* and *B* are sometimes given too readily" (Goodwin, 2011, p. 80). The concept of grade inflation has been with us ever since. Grade inflation is derived from the belief that rigor equals a scarcity of high grades and that the purpose of grading is to sort and rank. Grade inflation is assumed to be present when grades or grade point averages go up without a similar rise in achievement (Kohn, 2002; Zirkel, 2007). The belief is that if there are too many high grades, then they must be less accurate or not deserved. But higher grades could indeed be accurate.

> 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, failure to test well, and failure to have any influence at all on

> the intellectual lives of students. (Milton, Pollio, & Eison, 1986, p. 209)
>
> This decades-old debate won't be resolved as long as it is framed in terms of the 'normal' or 'bell' curve. Making this the first premise for grading students reflects a serious misunderstanding of this statistical concept. (Kelly, 2009, p. 696)

In the days before differentiation, when all students were taught one way and given the same amount of time to learn, this variation in achievement was easily accomplished. Teachers began to take pride in the ability of a test to discriminate and yield a bell-shaped distribution of grades. If necessary, conditions could easily be contrived to create a bell curve. There was no trick to giving bad grades—any teacher could make a test so hard that only a few students got *A*s. In an effort to make learning more rigorous, teachers often just created *more* work. But often that work was not more intellectually challenging. The *more is always better* argument ignored the quality of work and level of learning required (Vatterott, 2007). Rigor and difficulty was often equated to the amount of work done by students rather than the complexity and challenge of the work (Williamson & Johnston, 1999).

Institutionalizing failure. Unfortunately, sorting and ranking learners became a pervasive practice and a most formidable obstacle to student success (Canady & Hotchkiss, 1989). In far too many schools, the bell-shaped curve became the icon of sorting and ranking—it dictated programs and created the tradition of norm-comparisons. The bell curve set up "unnecessary and counterproductive scarcities of student success in competitive, win-lose environments" (Bonstingl, 1992, p. 7).

Use of the bell curve led to an acceptance of student failure—educators began to make plans for some students to fail. Schools established policies for retaining elementary students and developed programs for alternative schools and credit recovery for secondary students. Such practices reinforced the belief that

some students could not learn and perpetuated a system that not only allowed for but actually expected failure. In the sorting and ranking system, when failure did occur, it was viewed as the fault of the student and the blame was placed squarely there. The belief was that there was something wrong with those students and they needed to be fixed. In many ways, sorting and ranking practices institutionalized failure and conveniently absolved teachers of the responsibility for student failure (Vatterott, 2007).

Belief #2: Not Everyone Deserves an *A*

Born from our roots as moral educators and our fondness for sorting and ranking, the concept of an *A* student is embedded in our culture as an icon of discipline, responsibility, and hard work. Many people feel strongly that grades should reflect more than learning. We view grades as a package deal; to succeed, students must have it all—academic achievement *and* moral virtues. An *A* student is one who is not only smart—he or she must be learned *and* virtuous. For many teachers, that's what an *A* student is.

This belief is played out in schools every year. Here are a few common scenarios:

Jimmy, a surly and difficult 7th grader, is regularly sent to the office for disrespect and fighting. He is often absent, yet manages to make good grades on in-class work and tests. After absences, he fails to contact the teacher to make up missing work. Due to zeros for missing work, he is receiving *D*s in several classes. It is May and time for retention decisions. In spite of evidence that he has mastered the 7th grade curriculum, his academic team of teachers wants to retain him in the 7th grade "because he doesn't *deserve* to be promoted to the 8th grade."

Zack, a popular high school senior, is considered an excellent student. He consistently earns *A*s and *B*s and recently received a scholarship to Cornell. It's the end of the school year and he has a take-home exam due for his literature class. The exam counts

40 percent of his final course grade. Even though the exam demonstrates his mastery of the content at an *A* level, because the assignment was turned in a day late, he receives a zero. Zack may fail the class and his scholarship at Cornell may be in jeopardy. Last year in a similar situation with the same teacher, a student's acceptance to college was rescinded based on the effect of this practice on his final grade point average.

Some teachers feel it is their duty to sort and rank by intellect and virtue, that they have an obligation to society to "separate the wheat from the chaff." But as the scenarios show, sometimes their efforts result in decisions that adversely affect students.

Belief #3: Grades Motivate Learners

The belief that grades motivate learners was a logical extension of educators' homage to behaviorism. This belief has evolved from a series of misconceptions. The first misconception is that learning is only a means to an end—to escape punishment or get a reward, that learning has no intrinsic value, and that students would not be interested in learning for its own sake. If learning is an inherently distasteful process, then students must be coerced and grades can be a method of coercion (Stiggins, 2005; Vatterott, 2007). How sad if this were truly the case.

The second misconception is that a single entity called motivation exists, that students either have it or don't have it, and it can be manipulated by external forces (behaviorism again!) (Glasser, 1992; Stiggins, 2005). In other words, it is possible to motivate someone to do something they don't want to do, given the right sort of carrots and sticks. The third misconception is that the most effective method is the use of rewards and/or punishment (as opposed to engagement, relevance, or meaningful tasks) and that grades are an effective reward and/or punishment for all students. Most teachers have come to realize that good grades motivate students who value good grades and bad grades

don't motivate anyone, with the possible exception of a few overachievers (Vatterott, 2007).

Unlike Pavlov's dogs, children are not primitive creatures who respond only to reward and punishment. They are complex thinking, feeling beings with a natural curiosity that we must be careful not to extinguish. And learning is not merely a behavior—it is a complex mental process that is mediated by many factors and is unique in each child.

Grades as Commodity

Our beliefs have led to an abuse of grades. What began as a way to sort and reflect the quality of learning has evolved into an obsession, the weapon of choice for teachers, and a prized commodity for students and parents. In the behaviorist tradition, children are bribed, punished, and rewarded for the letters on the report card. Children and parents worship grades as the measure of a student's success. To them, grades have come to represent the relative value of a learner (regardless of the criteria used to arrive at the grade).

To students, grades have come to represent how hard they worked and how well they followed the rules. Students are quick to protest a grade that might actually reflect learning, if it is incongruent with their idea of what a grade means. For one student, the complaint went like this: "I attended every class and demonstrated an exemplary amount of participation. I was under the impression that I would earn an *A* with the effort I had applied." *Under the impression* is exactly our problem. Students have come to believe that effort (however weak), not learning, earns them the *A*. To parents, good grades reassure them that their child is a smart and successful student. We intensify the value of the grade by awarding honor roll status to the top grade-getters and affix stickers to our cars that state "My child is an honor roll student at Meadowbrook Elementary School." In response, other cars sport

bumper stickers that say "My kid can beat up your honor student" or "My dog is smarter than your honor student."

Truth be told, we like our norm-based comparisons. As a competitive culture, we *like* winners and losers and the fact that there is only so much room at the top. It is a way to demonstrate (we believe) that school is a meritocracy—that hard work is rewarded, and that a hierarchy of achievement exists. This system secures the fate of those few at the top, who are then given access to the best high schools and colleges. So the tradition of honor rolls, class rank, and valedictorians lives on.

Grades do reflect a meritocracy, but it's the wrong meritocracy. What is the merit? Is it learning, completion of all assigned tasks, compliance, or responsibility? A true meritocracy based only on learning would require an attitudinal shift. It would require us to let go of those other qualities we value in students, or at least agree to count them differently.

In our relentless pursuit of the almighty *A* and the perfect GPA, something got lost—learning. Grades became the be-all and end-all, the goal itself, not an indicator of achieving the goal of learning. Grades have become the commodity, the badge of success and smart, the ticket to college. But what do they really mean?

Forces Driving Change

Grades are often not an authentic reflection of learning. The way we currently use grades contributes to other problems in education. If changing grading practices could precipitate broader changes in teaching and learning, it's possible that our mediocre academic standing in the world could be greatly improved. Now is the time for change. Cultural and educational forces have created the perfect storm for changing grading practices. These forces compel us to action. These forces provide us with a much needed reality check about grades.

Reality Check #1: No Child Left Behind

Our earliest reality check came when No Child Left Behind (NCLB) became part of our world. With NCLB came accountability. Oh, wait—we need to worry about what our students *learn*, not just how well they *obey*? We can no longer simply fail those who don't learn and move on? "All must be proficient," NCLB said. This was a foreign concept to teachers—we had never been expected to ensure that all students were proficient. We didn't know *how* to do that. We were not even sure that it was possible. There were *always* some that failed. Wasn't that what rigor meant? We were accustomed to institutionalized failure. Then when we attempted to implement the concept of "no child left behind," we did it within the old sort and rank system. We taught them one way and didn't know how to differentiate, or if it was even fair to differentiate. We were stuck in a mindset of the past—that told us to sort and rank and to create a bell-shaped curve; a mindset that told us that success was supposed to be rare, not common; a mindset that told us that some were destined to fail.

The movement for learning proficiency revealed that what we thought grades meant was an illusion. If we accepted that standardized tests, flawed as they were, did reflect *some* measure of student learning, then the mismatch between the grades students were receiving and *actual* learning was stark. NCLB exposed a dirty little secret—*grades don't equate with performance on standardized tests*. Some students who were compliant and hard workers got good grades but did poorly on the standardized tests. Some students tested well but received poor grades because of missing assignments, late work, or bad behavior.

We came to realize that often grades didn't reflect proficiency in learning at all. With so many factors other than learning clouding the grade, it was hard to say what the grade meant. For students it became a chess game—a game of moves and strategies to get the grade. The further students progressed in school, the better they became at the game; the better they became at "doing

school" (Pope, 2001). By high school, students were quite adept at manipulating the system, cheating, and taking short cuts, all to get the grade.

> An *A* grade, therefore, did not necessarily mean that the students learned and retained content area knowledge and skills or that they understood important concepts or theories: rather, the grades proved that the students were adept at providing the teachers with the information required on tests and quizzes, and that they had memorized these facts and figures (or copied them from peers) just long enough to "ace" the exams and then move on to the next set of tasks. (Pope, 2001, p. 156)

Accountability for learning demands grades that are reflective of learning.

Reality Check #2: Grades Are Misleading About Succeeding

The implicit promise of good grades goes like this—if you work hard and get good grades, then you'll be successful in life. Good grades are viewed as a marker of success and responsibility, and our culture routinely rewards them as such. Parents pay cold hard cash, schools give away baseball tickets, and local businesses routinely give discounts for good grades. A puzzling example is that good grades in high school earn students cheaper car insurance. Why—because good students are safer drivers or because good grades mean you are an accomplished rule-follower who will follow the rules of the road?

Parents and students obsess about grades as the ticket to a better future. But do good grades correlate with future academic success? Yes and no. First, the good news. Yes, good grades in high school do correlate with ACT scores and first year college grades. There is a clear and consistent relationship between high school grade point average and scores on the ACT—as the score on the ACT rises, so does the average high school grade point average for that score (ACT, 2005). There is, however, evidence that high

school GPAs have been rising at a proportionately higher rate than scores on the ACT (Zirkel, 2007). One indicator is that the average high school GPA associated with specific scores on the ACT has risen. From 1991 to 2003, the average high school GPA of students receiving a composite score of 20 on the ACT rose from 2.85 to 3.10. For students earning a composite score of 28, the average high school GPA rose from 3.5 to 3.65 (ACT, 2005).

> Between 1991 and 2003, the mathematics grades of high school students taking the ACT exam rose from a grade point average of 2.80 to 3.04, whereas their average score on the math portion *[of the ACT]* rose only slightly, from 20.04 to 20.55 on a 36 point scale. (Goodwin, 2011, p. 80)

A similar relationship was shown in ACT English scores when compared to high school English grade point averages. In another comparison, although student high school grades in reading rose between 1992 and 2005, student scores on the reading portion of the National Assessment of Educational Progress (NAEP) declined (Schmidt, 2007).

What about the correlation between high school grades and college grades? Both high school grade point average and ACT scores correlate with college grades and successful completion of the freshman year (Beecher & Fischer, 1999; Geiser & Santelices, 2007). "ACT and high school grade point average seem to be effective predictors of college success regardless of course content because they are indicators of generic abilities and motivation" (Beecher & Fischer, 1999, p. 4). In other words, the relationship may exist because students have learned how to get good grades, not necessarily because they took more challenging courses in high school.

Now for the bad news. Grades are misleading when it comes to predicting success because high grades may help students *get into* college, but they don't necessarily prepare them academically to *succeed* in college. The reality is that in the United States,

30 percent of freshmen at four-year institutions will drop out (Goodwin, 2011) and only 54 percent of students entering college will complete a degree (Stewart, 2012). Interestingly, when college graduation rates are correlated with high school grade point averages, students with GPAs below 3.25 have graduation rates at 44 percent or lower. Only students with GPAs above 3.25 show graduation rates of above 50 percent (OIRP, 2011). The higher the high school GPA, the better predictor it is for college graduation (Stumpf & Stanley, 2002). But a college graduation rate of 54 percent is still a dismal figure.

High school students, busy chasing the grade, often arrive at college without a solid academic foundation to succeed in college classes. "Too many students now enter with advanced courses on their resumes but little grasp of the all-important basics" (Oxtoby, 2007, p. 43).

When looking at a high school transcript, it is impossible to know whether a *B* in high school chemistry indicates the student understood 80 percent of the concepts or that the student understood little and padded poor test grades with copied homework and extra credit.

> About a third of first-year students entering college had taken at least one remedial course for reading, writing, or math. The number is even higher for black and Hispanic students. At public two-year colleges, that average number rises to above 40 percent. And having to take just one remedial course is highly correlated with failure to graduate from college. (Friedman & Mandelbaum, 2012, p. 116)

By focusing on the high grade point average for college admission, some students have been shortchanged in learning, inadvertently making them *less* prepared academically while making them very talented at working the system. We thought that we were rewarding the right thing—completion of tasks, compliance, promptness. But in that process if we devalued mastery of deep

conceptual learning, we have hampered students' future success. Maybe the grading practices that we *thought* were preparing students for the future really weren't.

Reality Check #3: The Common Core State Standards Change Everything

The initiative for the Common Core State Standards originated with the nation's governors and education commissioners, through their representative organizations, the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO). At this writing, 45 states have agreed to implement these standards (www.corestandards.org). The goal of the standards is to prepare students for the future, to ensure that all students are college and career ready, and that they are able to compete in the global economy. The standards attempt to raise the bar for what students know and are able to do. The standards go beyond previous standards in that they require students to use and apply complex knowledge and skills to solve unfamiliar problems. This level of application of knowledge and skills is one area in which U.S. students are particularly lacking. For instance, on international tests, U.S. students fared relatively better on the TIMMS in part because the multiple choice questions asked them to reproduce curriculum content. They did relatively poorly on PISA, which required them to apply knowledge to real-life situations (Stewart, 2012). Those results exposed another dirty little secret in U.S. education. Although standards and standardized tests have supposedly driven instruction for years, we now see that *we have been focusing too much on low level rote learning.*

Too often, we have neither allowed nor expected students to think. We have filled their heads with facts and formulas and rewarded them for reciting it. *We* have done the analyzing, synthesizing, and evaluating instead of expecting our students to do it. *We* have done too much of the work of learning, perhaps because

we didn't trust them to *want* to do the work, or perhaps because we weren't sure they were *able* to do the work. This failure to require higher level thinking is part of the reason our students are not prepared for college. (After assigning a synthesis level task to my college sophomores, I was dismayed to have a student ask, "Where do I find the answer?" "You have to create it!" I replied.)

Putting it bluntly, our definitions of what it means to teach and what it means to learn are outdated and simplistic. Our expectations have been seriously out of sync with what the future will demand of our students. The implementation of the standards should change that. They will affect not only what we teach, but how we teach, and how we assess what students have learned. (The effects of standards will be discussed more fully in Chapters 3 and 4.) To successfully navigate the standards, student grades will need to reflect mastery of skills, not memory of content. This goal is congruent with the skills students will need to succeed in a new world.

Preparing Students for the New World

The world we are preparing our students for has changed. In the past we were preparing them for an industrial world and top-down management—obey, meet deadlines, follow rigid rules, punch the time clock. Today we must prepare them for a world in which they must know how to take initiative, self-advocate, solve problems, be creative, and accomplish tasks without minute to minute supervision. Tomorrow's jobs will require critical thinking, sophisticated communication skills, handling non-routine complex tasks, and working collaboratively with others to solve problems (Friedman & Mandelbaum, 2011; Stewart, 2012). As the world has changed, the outcomes we want for education must change. We need students to be both college and career ready. What skills and dispositions are needed? The standards have defined them. Students who are college and career ready are as follows:

- motivated to learn independently of external rewards and punishments.
- self-directed learners who know how to assess their own learning needs.
- inclined to seek out and use resources to assist them in learning.
- exhibit a willingness to try, persistence, and a belief that effort will pay off in eventual success (www.corestandards.org).

Many traditional educational practices, such as rote learning and the use of grades as reward and punishment, have interfered with, if not prevented, the development of these essential skills and dispositions in students. We must reexamine our teaching, learning, and assessment practices to decide which type of student we want to produce.

- Do we want students who can memorize and repeat or students who can analyze, synthesize, and problem solve?
- Do we want students who are excited and engaged and involved in their learning or students who obediently slog through whatever tasks they are given?
- Do we want students whose goal is to get the grade at any cost or to find meaning in what they do?

The kind of students that we produce will determine the kind of adults who will inhabit the United States and our global community. Chapter 2 will examine in detail how traditional grading practices hinder the development of the skills our students need for a successful future.

References

ACT, Inc. (2005). Are high school grades inflated? *Issues in college readiness.* Author. ERIC: ED510537.

Beecher, M., & Fischer, L. (1999). High school courses and scores as predictors of college success. *Journal of College Admissions*, Spring/Summer issue 163, 4–9.

Bonstingl, J. J. (1992). The quality revolution. *Educational Leadership*, *50*(3), 4–9.

Brookhart, S. M. (2008). *How to give effective feedback to your students*. Alexandria, VA: ASCD.

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership*, *69*(3), 10–14.

Brooks, J. G., & Dietz, M. E. (2012). The dangers and opportunities of the Common Core. *Educational Leadership*, *70*(4), 64–67.

Canady, R. (2010). *School policies and grading practices that increase/decrease the odds for student success*. National Middle School Association's annual conference, Baltimore, MD.

Canady, R. L., & Hotchkiss, P. R. (1989). It's a good score! Just a bad grade. *Phi Delta Kappan*, *1*(1), 68–71.

Canter, L. (1976). *Assertive discipline: A take-charge approach for today's educator*. Los Angeles: Lee Canter and Associates.

Canter, L., & Canter, M. (1992). *Assertive discipline: Positive behavior management for today's classroom*. Santa Monica, CA: Lee Canter and Associates.

Chappuis, J. (2009). *Seven strategies of assessment for learning*. Portland, OR: ETS Assessment Training Institute.

Curreton, L.W. (1971). The history of grading practices. *National Council on Measurement in Education*, 2 (4), 1–9.

DeLorenzo, R. A., Battino, W. J., Schreiber, R. M., & Gaddy Carrio, B. (2009). *Delivering on the promise: The education revolution.* Bloomington, IN: Solution Tree.

Durm, M. W. (1993). An *A* is not an *A*: A history of grading. *The Educational Forum*, 57 (Spring), 1–4.

Dweck, C. (2007). *Mindset: The new psychology of success*. New York: Ballantine.

Education Week. (2013). Diplomas count 2013, executive summary: Making progress and trying again. *Education Week*, *32*(34), 1.

Freiberg, H. J. (1999). Beyond behaviorism. In Freiberg, H. J. (Ed.), *Beyond behaviorism: Changing the classroom management paradigm* (pp. 3–20). Needham Heights, MA: Allyn and Bacon.

Friedman, T. L., & Mandelbaum, M. (2012). *That used to be us: How America fell behind in the world it invented and how we can come back.* New York: Picador.

Geiser, S., & Santelices, M. V. (2007). Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. *Center for Studies in Higher Education Research and occasional paper series: CSHE.6-07*. Berkeley, CA: University of California, Berkeley.

Glasser, W. (1992). *The quality school: Managing students without coercion*. New York: HarperCollins.

Goodwin, B. (2011). Grade inflation: Killing with kindness? *Educational Leadership*, *69*(3), 80–81.

Guskey, T. R. (1996). *Communicating student learning. ASCD yearbook*. Alexandria, VA: ASCD.

Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan*, 87(9), 670–675.

Guskey, T. R. (2007). Using assessments to improve teaching *and* learning. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 15–29). Bloomington, IN: Solution Tree.

Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership*, *69*(3), 16–21.

Guskey, T. R. (2015). *On your mark: Challenging the conventions of grades and reporting.* Bloomington, IN: Solution Tree.

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning.* Thousand Oaks, CA: Corwin.

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

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relative to achievement.* New York: Routledge.

Herman, J., & Linn, R. (2014). New assessments, new rigor. *Educational Leadership*, *71*(6), 34–37.

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.

Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.

Kelly, T. F. (2009). Grade inflation: Sense and nonsense. *Phi Delta Kappan*, *90*(9), 696.

Kirschenbaum, H., Simon, S. B., & Napier, R. W. (1971). *Wad-ja-get: The grading game in American education*. New York: Hart Publishing.

Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes.* New York: Houghton Mifflin.

Kohn, A. (1999). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes (2nd edition).* New York: Houghton Mifflin.

Kohn, A. (2002). The dangerous myth of grade inflation. *The Chronicle of Higher Education*, *49*(11), B7.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD.

McClellan, B. E. (1999). *Moral education in America: Schools and the shaping of character from colonial times to the present.* New York: Teachers College Press.

Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades*. San Francisco: Jossey-Bass.

Moss, C. M., & Brookhart, S. M. (2012). *Learning targets: Helping students aim for understanding in today's lesson.* Alexandria, VA: ASCD.

O'Connor, K. (2009). *How to grade for learning K–12 (3rd edition).* Thousand Oaks, CA: Corwin.

OIRP (2011). *High school GPA and SAT score effects on graduation rates*. Office of Institutional Research and Policy Studies, March 18, 2011.

Oxtoby, D. W. (2007). The rush to take more AP courses hurts students, high schools, and colleges. *Education Digest*, *73*(1), 43–46.

Pope, D. C. (2001). *'Doing school': How we are creating a generation of stressed out, materialistic, and miseducated students.* New Haven, CT: Yale University.

Reeves, D. (2011). *Elements of grading: A guide to effective practice.* Bloomington, IN: Solution Tree.

Reeves, D. B. (2008). Leading to change/Effective grading practices. *Educational Leadership*, *65*(5), 85–87.

Schimmer, T. (2012). *Ten things that matter from assessment to grading*. Toronto, Ontario: Pearson.

Schimmer, T. (2013a). *Infused assessment*. Presented at the Assessment Training Institute, Portland, Oregon, July 8, 2013.

Schimmer, T. (2013b). *Accurate grading with a standards-based mindset*. Sound Grading Practices Conference: Leading the change in classroom grading. December 6, 2013. Portland, Oregon.

Schmidt, P. (2007, March 9). High-school students aim higher without learning more, federal studies find. *Chronicle of Higher Education*, *53*(27), A32.

Seligman, M. (1998). *Learned optimism*. New York: Pocket Books.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high school work in English. *School Review*, 20.

Starch, D., & Elliott, E. C. (1913). Reliability of grading work in mathematics. *School Review*, 21.

Stewart, V. (2012). *A world-class education: Learning from international models of excellence and innovation.* Alexandria, VA: ASCD.

Stiggins, R. (2007). Assessment *for* learning: An essential foundation of productive instruction. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 59–78). Bloomington, IN: Solution Tree.

Stiggins, R. J. (2005). *Student-involved assessment for learning.* Upper Saddle River, NJ: Pearson.

Strom, P. S., & Strom, R. D. (2007). Cheating in middle school and high school. *Educational Forum*, *71*(2), 104–116.

Stumpf, H., & Stanley, J. C. (2002). Group data on high school grade point average and scores on academic aptitude tests as predictors of institutional graduation rates. *Educational and Psychological Measurement*, *62*(6), 1042–1052.

Thayer, V. T. (1965). *Formative ideas in American education from the colonial period to the present.* New York: Dodd, Mead, and Co.

Tomlinson, C. (2008). The goals of differentiation. *Educational Leadership*, *66*(3), 26–30.

Tomlinson, C. (2014). The bridge between today's lesson and tomorrow's. *Educational Leadership*, *71*(6), 10–14.

Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character*. New York: Houghton Mifflin Harcourt.

Trible, L. (2012). New school grade system yanked for this year. *The Daytona Beach News-Journal,* October 4, 2012.

U.S. Department of Education. (2010). *Forest Grove High School (Transformational Model)*. Author: Washington, DC. Retrieved from http://www.ed.gov/oese-news/forest-grove.

U.S. Department of Education. (2012a). Science Literacy: Average Scores (Overall). Retrieved from https://nces.ed.gov/surveys/pisa/pisa2012/pisa2012highlights_4a.asp

U.S. Department of Education. (2012b). Mathematics Literacy: Average Scores (Overall). Retrieved from https://nces.ed.gov/surveys/pisa/pisa2012/pisa2012highlights_3a.asp

U.S. Department of Education (2012c). Reading Literacy: Average Scores (Overall). Retrieved from https://nces.ed.gov/surveys/pisa/pisa2012/pisa2012highlights_5a.asp

Varlas, L. (2013). How we got grading wrong, and what to do about it. *Education Update*, *55*(10), 1, 6–7.

Vatterott, C. (2007). *Becoming a middle level teacher: Student focused teaching of early adolescents.* New York: McGraw-Hill.

Vatterott, C. (2009). *Rethinking homework: Best practices that support diverse needs.* Alexandria, VA: ASCD.

Vatterott, C. (2010). Five hallmarks of good homework. *Educational Leadership, 68*(1), 10–15.

Vatterott, C. (2011). Making homework central to learning. *Educational Leadership*, *69*(3), 60–64.

Vatterott, C. (2014). Student-owned homework. *Educational Leadership*, *71*(6), 39–42.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (expanded 2nd edition). Alexandria, VA: ASCD.

Wiliam, D. (2007). Content *then* process: Teacher learning communities in the service of formative assessment. In D. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 182–204). Bloomington, IN: Solution Tree.

Williamson, R., & Johnston, J. H. (1999). Challenging orthodoxy: An emerging agenda for middle level reform. *Middle School Journal*, *30*(4), 10–17.

Zirkel, P. A. (2007). Grade inflation: High schools' skeleton in the closet. *Education Week*, *26*(29), 40, 30.

Zmuda, A. (2008). Springing into active learning. *Educational Leadership*, *66*(3), 38–42.

About the Author

Cathy Vatterott, PhD, is a professor of education at the University of Missouri–St. Louis, and was a middle school and high school teacher and a former middle school principal. She is the author of numerous articles about education and three books, *Academic Success Through Empowering Students* (NMSA), *Becoming a Middle Level Teacher: Student-focused Teaching of Early Adolescents* (McGraw-Hill) and *Rethinking Homework: Best Practices That Support Diverse Needs* (ASCD). Dr. Vatterott is considered a national expert on the topic of K–12 homework and is often interviewed by reporters from the *New York Times, National Public Radio, BBC World News Service*, and *USA Today*, as well as writers from parenting magazines and educational organizations.

Discussions about the grading of homework in her presentations in the United States and Canada revealed a major disconnect between how grades are typically used in K–12 classrooms and our goal of helping students meet academic standards. These discussions have been the catalyst for her latest research about K–12 standards-based grading. She can be reached at vatterott@umsl.edu or through her website at www.homeworklady.com.