



## Douglas Reeves

A school committed to high levels of learning for all students must confront these questions: "What are the most essential skills and knowledge each student must acquire, and what is the standard they must achieve to demonstrate their proficiency?" No writer in America has been more effective in presenting a persuasive rationale for the ethical imperative of articulating standards of learning for all students than Dr. Douglas Reeves.

Dr. Reeves is founder and chairman of the Center for Performance Assessment, an international organization dedicated to improving student achievement and educational equity. He also serves as a faculty member for leadership programs sponsored by the Harvard Graduate School of Education. A prolific writer, Dr. Reeves has authored many articles and books, including *Assessing Educational Leaders: Evaluating Performance for Improved Individual and Organizational Results*; *Accountability for Learning: How Teachers and School Leaders Can Take Charge*; *The Leader's Guide to Standards: A Blueprint for Educational Excellence and Equity*; and the best-selling *Making Standards Work*, now in its third edition.

In this chapter, Dr. Reeves advances three main arguments and offers practical and constructive approaches to standards, assessment, and accountability. He contends that schools and school systems must translate standards into a set of rational, relevant, and above all *focused* expectations or "power standards"; those standards must be accompanied by frequent, common assessments in the classroom; and state and local accountability systems must include not only test scores, but explicit indicators of adult behavior. A common thread that runs through his work is the need for the collaborative culture of a professional learning community if educators are to meet the challenges they face.

Dr. Douglas Reeves is a frequent keynote speaker in the United States and abroad for education, government, and business organizations. He can be reached at [dreeves@makingstandardswork.com](mailto:dreeves@makingstandardswork.com).

On Common Ground: The Power of Professional Learning Communities  
Chapter 3, Doug Reeves, 2005, National Educational Service

### Chapter 3

## Putting It All Together: Standards, Assessment, and Accountability in Successful Professional Learning Communities

Douglas Reeves

In an astonishingly short period of time, the standards movement has swept the nation. While only a handful of states had adopted academic standards in the early 1990s, the use of standards is now a matter of federal law and all 50 states have adopted one version or another. Although the terminology surrounding standards varies widely, the notion that an educational system should have a coherent set of expectations about what students should know and be able to do is widely held in public and private schools throughout the world. As standards have become commonplace in the United States, forests have been cleared to publish the documents accompanying academic standards. Every textbook and curriculum document in the land claims to be "standards-based," as if such an imprimatur

were the educational equivalent of a “Good Housekeeping Seal of Approval” of a bygone era. The cynics were certain that—along with Outcomes-Based Education, Behavioral Objectives, Mastery Learning, and a host of other reform ideas—standards would go the way of the dinosaur and eventually be of interest only to educational paleontologists of a future era. Standards advocates, on the other hand, were certain that the battle was won with the passage of the No Child Left Behind Act in 2001. Both sides were wrong. The one thing that can be said with certainty is this: Standards are not enough.

This chapter advances three arguments:

- First, standards in their present form are inadequate as a foundation for improved achievement and professional practices. Schools and school systems must translate standards into a set of rational, relevant, and above all *focused* expectations that I have labeled “power standards.”
- Second, standards must be accompanied by frequent common assessments in the classroom. While the nation may be, according to the charges of many critics, over-tested, our students are actually under-assessed. This critical distinction lies at the heart of effective reform efforts.
- Third, state and local accountability systems must include not only test scores, but also explicit indicators of adult behavior such as teaching practices, curriculum, leadership, and other influences on student achievement.

These three arguments are as important for senior leaders and policymakers at the district, state, and national level as they are for classroom educators. Without the support of each of these three pillars—standards, assessment, and accountability—the roof of educational progress will sag, crack, and crumble. As the last decade has taught us, standards without assessment are fantasies. Assessments not linked to standards are prescriptions for frustration. Accountability systems that evaluate students without taking into account the actions of educators and leaders are as useful as the health club whose only piece of equipment is a scale.

The members of this health club weigh themselves diligently and frequently, but are oblivious to the relationship between those numbers and what a more thoughtfully constructed health program might include: diet, exercise, and careful attention to the individual medical needs of each person. When the health club only measures weight, no one knows if the apparently successful weight loss is the result of diet and exercise or anorexia and drug abuse. If the “score”—weight or test scores—is the only information that is relevant, neither the organization nor the individuals within it have the opportunity to make systematic improvements.

The framework of a professional learning community (Dufour, DuFour, Eaker, & Karhanek, 2004) is inextricably linked to the effective integration of standards, assessment, and accountability. The concept of professionals in community with one another flies in the face of the line guaranteed to garner applause at many gatherings: “Just leave me alone and let me teach!” Teaching has traditionally been a solitary enterprise, with idiosyncratic judgment and personal preference

trumping external demands for consistency, fairness, and effectiveness. While few people would dine in a restaurant where the chef tells the city health department to “Just leave me alone and let me cook!” or place our medical care in the hands of a physician who resisted external accountability and standards of professional practice, we regularly place our children in classrooms where collaboration is not only absent, but also actively resisted. Recognizing that organizational culture and structure will influence behavior, the leaders of professional learning communities balance the desire for professional autonomy with the fundamental principles and values that drive collaboration and mutual accountability.

### **Power Standards: From Fantasy to Focus**

While academic standards vary widely in their specificity and clarity, they almost all have one thing in common: there are too many of them (Marzano & Kendall, 1998). It is not uncommon to find standards and accompanying curriculum documents that specify 80 or more learning objectives for a single subject in a single grade. Only a few states have prioritized standards, thus leaving the majority of teachers with two options: curriculum by default or curriculum by design. The default option stems from the fantasy that if an expectation is approved by a state department of education and supported with three-ring binders created by a district curriculum department, then it will happen in the classroom.

The strength of one’s belief in the standards fantasy is inversely proportional to the distance one is from the classroom. While many classroom teachers who face diverse learning needs and limited time know that rapid coverage of standards is not a substitute for student understanding, standards and curriculum

designers at the district, state, or national levels continue to publish documents that are distinguished more by their girth than their effectiveness.

Curriculum by default is the result of the urgency with which we often act—we inevitably run out of time, energy, and patience by the end of the school year. Curriculum by design, however, reflects decisions made before the school year begins. Curriculum by default involves exasperation and resignation; curriculum by design involves prioritization and planning. Later in this chapter, we will consider how to use the concept of “power standards” to supplant curriculum by default with thoughtful design.

### **Standards: The Best Alternative to the Bell Curve**

Unfortunately, the response of some people who have reasonable criticisms of bad standards is the unwarranted conclusion that state standards in general are flawed and should be discarded (Kohn, 1999; Ohanian, 1999). A better approach is to recognize that while standards are deeply flawed, they can and must be improved. If we fail to take such a constructive approach, then the inevitable result will be a return to the bell curve, a process of evaluating student achievement not based upon objective comparison of student performance to a standard, but by the pernicious and destructive process of evaluating student performance based upon who beat whom (Reeves, 2001b).

However flawed standards may be in their present form, they are vastly superior to the bell curve. The comparative approach to evaluation represented by the bell curve not only devastates underperforming students but also provides a sense of inappropriate complacency to the student who performs



well relative to fellow students, but who nevertheless does not meet an objective standard. Every reader can think of a student who scored “above average” on a national norm-referenced literacy assessment, but who was unable to meet a state or local standard for writing and reading comprehension.

### Criteria for Power Standards

The answer, therefore, is neither the continued proliferation of fantasy standards nor the rejection of all standards, but the establishment of power standards (Ainsworth, 2003; Reeves, 2000). In order to be identified as a power standard, an academic expectation (call them what you will—standards, objectives, curriculum elements, benchmarks) must meet three criteria:

1. Endurance
2. Leverage
3. Essential for the next level of instruction

First, it must have **endurance**, a characteristic of those standards whose importance lasts longer than a few nanoseconds after the termination of a state test. The principle of endurance is reflected in the recurring nature of key skills and knowledge that students must display. In some cases, enduring standards can be found in standards documents themselves. Reading comprehension, descriptive writing, and inferential reasoning can be found in academic standards from the early primary to the final secondary years. In other cases, however, the principle of endurance occurs in implicit standards, such as those involving time management, project management, and personal organization. While state academic content standards rarely consider these matters, every teacher who has considered

the fate of a failing student knows the relative importance of academic content and time/project management, and those teachers will choose the latter as an enduring priority for student success.

Second, a power standard must have **leverage**. The principle of leverage means that success in one standard is very likely to be associated with success in other standards. The most notable application of the principle of leverage is the association between nonfiction writing, a power standard if ever there was one, and student success in math, social studies, and science (Calkins, 1994; Darling-Hammond, 1997; Reeves, 2000, 2001a, 2004a, 2004b). Another example of the principle of leverage is the mathematics standard requiring students to create and draw inferences from tables, charts, and graphs. This standard will not only lead to student success in mathematics, but will also help students record and interpret results from a science experiment and better analyze and understand data in economics, geography, history, health, and physical education courses.

The third criterion for power standards is that it is **essential for the next level of instruction**. When I ask audiences of educators what they are willing to give up from their curriculum, they are almost invariably silent, insisting that every element of their curriculum is vital, required by state standards, might be mentioned on a state test, and will be needed by their students in future years. But when I ask the same group to give advice to teachers in the next lower grade about what students must know and be able to do in order to advance to the next class with success and confidence, I have never—not a single time in thousands of cases—heard the words, “The teacher in



the grade lower than me should cover every single standard.” Rather, when giving advice to colleagues in a lower grade, educators are remarkably brief and balanced. They are able to give a list of a dozen or fewer critical educational, behavioral, and organizational needs that every student must meet. They routinely disregard the presumption that teacher coverage of every standard is a relevant part of the conversation.

### **The Role of Professional Learning Communities**

While supposed respect is granted by various constituencies with the inclusion in the standards of a favorite court case, geographic feature, constitutional amendment, or historical event, real respect for these subjects is diluted by the constrictions of time and the imminence of state-mandated tests.

Standards, in sum, are a necessary but insufficient element of the equation for improving student achievement and educational equity. Without the articulation of standards, the evaluative scheme for students will inevitably revert to the bell curve, discouraging underperforming students and failing to challenge high-performing students. With only the articulation of standards, however, students and teachers are left with unprioritized curriculum chaos. Professional learning communities add value to standards not by merely delivering them to the schoolhouse door, but by also analyzing, synthesizing, and prioritizing them in a way that allows every teacher to wisely allocate time and instructional focus. Only when these steps have been taken can a school proceed to the second component necessary for excellence: assessment for learning.

### **Assessment for Learning: The Key to Continuous Improvement**

Stiggins and associates (2000, 2004) and Wiggins (1998) thoughtfully distinguish between “assessment of learning” and “assessment for learning.” While the former is designed merely to render a report, the latter is designed to actively engage students and teachers and improve their performance. With far less literary elegance, I have used the analogy of physicals and autopsies (Reeves, 2000, 2004a).

Without putting too fine a point on the metaphor, physicals at a certain point in life can be an uncomfortable ordeal but, on the whole, they are preferable to and less intrusive than autopsies. The wise physician does not use the annual physical only to evaluate the patient, but also to recommend improvements in lifestyle. From the best of our family doctors, we receive not the hieroglyphics of lab results, but also candid advice to replace candy with carrots and the television with a treadmill. The keys to assessment for learning—the physical rather than the autopsy—are consistency, timeliness, and differentiation.

### **Consistency in Assessment**

As a teacher through and through, I cannot resist giving assignments, even when my students are readers on the other side of a printed page. I have learned that the best assignments are those that provide opportunities for discovery, challenge, and genuine “A-ha!” moments. Please consider the following assignment, as it will make the point more vividly than any external evidence or research analysis that I could otherwise provide.

1. Go to three different classrooms in the same grade that are studying the same subject and learning the same unit.



2. From each classroom, gather three pieces of student work that have received a grade of B or, if the teacher uses the language of standards, a grade of “proficient.” This is not the best work or the worst, but clearly work that the teacher and students regard as acceptable. You now have nine pieces of student work, all of which purport to represent student success in the same subject, same grade, and same unit.
3. Cut off the name of the student, school, and teacher. Blot out any indicator of the grade on the paper. You are thus left with nine pieces of otherwise undifferentiated student work.
4. Copy these nine exhibits and distribute them at a faculty meeting, telling your colleagues only that these nine pieces of work are all from the same grade and represent student work on the same subject.
5. Ask the following penetrating questions: What do you think? What do you notice? You can expect to hear a cacophony of responses:
  - “What? I can’t believe that this is fifth grade work! My students do better than this!”
  - “Check out sample #4—it’s superior—are you sure that this student is in the right grade?”
  - “Somebody said that this work was okay? In my class, it would have flunked.”
  - “This is exceptional! I’d love to have students who took an assignment this seriously.”

Listen patiently to the observations of your colleagues.

6. Write their observations of dramatic differences in the student work samples on chart paper that will be visible to everyone in the room.
7. When they have exhausted their analysis of profound differences in the nine samples of student work, deliver your punch line:
 

“Colleagues, I am confused. All nine of these samples of student work were in the same subject, the same grade, and the same unit. All nine of them received the same grade. Yet all I have heard from you are insightful observations about the profound differences in these work samples. I really trust the accuracy of your observations, so that must mean that our individual evaluation of student work is woefully inconsistent. If assessment of student work is so unclear to us as college-educated professionals, how unclear and confusing must it be to our students?”

### **Timeliness: The “Nintendo Effect”**

If assessment feedback is to be effective, it is not enough that it is accurate and consistent. It must also be provided to the student in a timely manner (Marzano, Pickering, & Pollock, 2001). A compelling metaphor for timely assessment feedback is what Jeff Howard, founder of the Efficacy Institute, refers to as the “Nintendo Effect.” Consider the distracted and inattentive student who, for reasons I cannot imagine, finds my middle school math class less than enthralling. This student appears to be constitutionally unable to sit still, focus, and concentrate. Before we reach for a pharmaceutical solution or

more commonly, give up on the student, perhaps we should observe the same student playing a Nintendo™ video game.

At the flip of a switch, the previously distracted and inattentive student is now silently focused like a laser and completely engaged. Even when he receives negative feedback (Nintendo players, after all, die at the conclusion of each game), he comes back for more, sometimes for hours at a time. What is Nintendo giving him that I am failing to provide in my math and writing classes? Feedback—but not just any feedback. Nintendo provides feedback that is specific, accurate, incremental, and timely. At the conclusion of every game, students have a very clear idea of what they did wrong and how to do better next time. How many students would play Nintendo if their scores were given to them days or weeks after playing the game? How many would play if they only received a score, but did not have a clear idea of how to improve their score? How many would play if they concluded that success at Nintendo was a function of skin color, language, or wealth? The cross-cultural success of Nintendo reminds us that timely feedback is effective for every single student. For the record, we might note that Nintendo provides the same accurate, incremental, and timely feedback to every student because Nintendo does not know the gender, ethnicity, or income level of the players. It just provides feedback that helps everyone improve.

I find the electronic game analogy troubling because, after all, school is much more important than a game, is it not? This is a good opportunity to ponder the seriousness that students and society in general give to games. Our society demands consistency and fairness in electronic games and all varieties of

sports. Watch closely the next time a close call by athletic officials goes against your team. Failures in consistency result in screams from the sidelines from parents, teachers, and students, all united in common cause. They all are providing collective feedback that is, to say the least, specific and timely. Their demand for consistency and fairness is fueled by indignation and zeal. How I wish that the emotional energy invested in games was also invested in academic performance and assessment. If the game analogy borders on the offensive for some readers, there are many other examples of effective feedback models in schools.

My favorite is the music teacher. When the cellos are out of tune or the altos are off-key, the music teacher does not haul out the grade book to document a reason for a low grade weeks hence. Great music teachers stop, correct the problem, and then improve individual, small group, and large group performance. Students leave every class knowing that they are better than when they came in. Lucy Calkins (1983, 1994), founder of the Columbia University Literacy Project, helps teachers create an environment for providing immediate and meaningful feedback on student writing in much the same way that great music teachers generate continued enthusiasm from their students. Of one of her students, Professor Calkins (1983) wrote the following when a group of visitors observed the child's superior writing:

"Susie is gifted," the visitors said, and I agreed. Gifted with the courage and the skills to experiment with writing. Gifted with the human potential to learn through trial and error.



The writing abilities of the children in this classroom were quite remarkable. Yet there had been nothing remarkable about their growth as writers. Susie's skills had not appeared full-blown out of nowhere nor had those of her classmates. Instead, the skills developed step-by-step, the way all learning happens. Susie's magic touch was not the result of a miracle but of experience and good teaching. She was not born with a talent for writing. She'd developed that talent through working at her craft. She'd gained a skill through the process of doing her best and then making her best better. (p. 88)

### **Differentiation**

The clarion call for consistency in previous paragraphs seems at first to fly in the face of the suggestion that effective assessments must be differentiated. Indeed, there are many people who have used the label of "differentiation" to justify terrible professional practices and low expectations. Effective differentiation in assessment does not dilute academic standards or rigorous expectations. Rather, differentiated assessment acknowledges that students must be able to show what they know in a variety of different ways. While some students can demonstrate proficiency using traditional multiple choice tests, others will demonstrate similar or higher levels of proficiency when they write, demonstrate, or otherwise engage in performances that display their knowledge and skills. The use of standards-based performance assessments is, in fact, more rigorous than the typical multiple-choice test (Reeves, 2002).

Teachers who want to more fully engage students in daily work, tests, and projects will find that they can maintain higher levels of student energy and commitment if they combine a consistently high level of academic rigor and expectations with a variety of alternative assessments. Rather than assuming that every student needs the same level of skill development, teachers create homework menus, engage the interest of students who might be bored by typical homework assignments, and help develop fundamental skills in students who came to the class without adequate preparation. Tests and projects can provide consistent levels of rigor and standards, while also providing a variety of alternative methods for demonstrating knowledge. An excellent way for students to demonstrate the depth of their understanding of a subject is for them to create their own test questions and write their own scoring rubrics (Ainsworth & Christinson, 1998), creating an optimal combination of engagement, rigor, and differentiation.

In the hands of a capable professional learning community, assessment is relentlessly constructive and focused on its singular purpose—the improvement of teaching and learning. Classroom assessments reflect consistently high expectations while providing a rich variety of methods for meeting a complex array of student needs. The feedback from these assessments is consistent—one teacher's "proficiency" is not another teacher's "superior"—and the communication of this feedback is timely. In the final section of this chapter, we will consider how to take the lessons of effective assessment and apply them to educational accountability systems involving many schools.





## Accountability: The Leadership and Learning Connection

Even the best standards and assessments will undermine student learning if professional practices are subverted by counterproductive accountability systems. Despite pervasive complaints that accountability policies of federal and state governments dominate school decision-making, there is a growing body of evidence that accountability can be constructive, comprehensive, and supportive of professional learning and student achievement (Reeves, 2000, 2004a; Schmoker, 2001). Effective accountability systems are not merely a litany of test scores; they also include three critical levels of information:

- **Tier 1** indicators reflect typical accountability data, including test scores and other data required for external accountability. But if we stopped there, as too many schools do, we would not understand the causes of improved achievement. It would be as if we knew an obese student had lost weight, but we did not know if the causes were diet and exercise or anorexia and drug abuse. That is why effective school accountability systems include Tier 2 data.
- **Tier 2** data are made up of measurable indicators that reflect professional practices in teaching, curriculum, and leadership.
- **Tier 3** of this accountability system is a school narrative, allowing teachers and school leaders to provide a qualitative context for quantitative data—the story behind the numbers.

Only with this system will accountability systems provide the appropriate context, allowing professional learning communities to explore both the causes and effects surrounding improved student performance and professional practices.

### Refined and Focused Standards

The three pillars of standards, assessment, and accountability support the effectiveness of professional learning communities. Remove one of those pillars and the structure will sag, crack, and eventually break. Standards must not merely be delivered from the state department to the schoolhouse door; they must be refined and focused. Assessment must not be the subject of annual academic post-mortems, but the focus of continuous discussion by professionals throughout the year. Accountability systems must focus not only on what students achieve, but also on how the adults in the system influenced that achievement.

### References

- Ainsworth, L. (2003). *Power standards: Identifying the standards that matter most*. Englewood, CO: Advanced Learning Press.
- Ainsworth, L., & Christinson, J. (1998). *Student-generated rubrics: An assessment model to help all students succeed*. Assessment Bookshelf Series. Orangeburg, NY: Dale Seymour Publications.
- Calkins, L. M. (1983). *Lessons from a child: On the teaching and learning of writing*. Portsmouth, NH: Heinemann.
- Calkins, L. M. (1994). *The art of teaching writing*. Portsmouth, NH: Heinemann.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.



- DuFour, R. (2002). The learning-centered principal. *Educational Leadership*, 59(8), 12-15.
- DuFour, R., DuFour, R., Eaker, R., & Karhanek, G. (2004). *Whatever it takes: How professional learning communities respond when kids don't learn*. Bloomington, IN: National Educational Service.
- Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional classrooms and "tougher standards."* NY: Houghton Mifflin.
- Marzano, R. J., & Kendall, J. S. (1998). *Implementing standards-based education (student assessment series)*. Washington, DC: National Educational Association.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Ohanian, S. (1999). *One size fits few: The folly of educational standards*. Portsmouth, NH: Heinemann.
- Reeves, D. B. (2000). *Accountability in action: A blueprint for learning organizations*. Denver: Advanced Learning Press.
- Reeves, D. B. (2001a). *101 questions & answers about standards, assessment, and accountability*. Denver: Advanced Learning Press.
- Reeves, D. B. (2001b, June 6). If you hate standards, learn to love the bell curve. *Education Week*, p. 48.
- Reeves, D. B. (2002). *Making standards work: How to implement standards-based performance assessments in the classroom, school, and district* (3rd ed.). Denver: Advanced Learning Press.
- Reeves, D. B. (2004a). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Reeves, D. B. (2004b). *101 more questions & answers about standards, assessment, and accountability*. Denver: Advanced Learning Press.

- Schmoker, M. (2001). *The results fieldbook: Practical strategies from dramatically improved schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Stiggins, R. J. (2000). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Stiggins, R. J., Arter, J., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Portland, OR: Assessment Training Institute.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

