

CHAPTER ONE

Accountability and High-Stakes Testing

Lee Shulman: *At the heart of my work on good teaching is the notion of a teacher as an enlightened, passionate intellectual.*

Carol Tell: *Do you find that this notion goes against the trend these days to measure the success of students—and their teachers—by standardized test scores?*

Lee Shulman: *The confusion stems from valuing standards, on the one hand, and embodying those standards in high-stakes assessments, on the other. The assessments end up corrupting the value of the standards. The standards get modified to be consistent with what we're able to measure in a high-stakes assessment. We have to ratchet down the standards and squeeze out all of the creative diversity because we want to be able to develop scoring keys that nobody can complain about or challenge.*

C. Tell (2001, pp. 6, 8)

2 STAYING ON TRACK

Standards, even when well implemented, can take us only part way to successful large-scale reforms. It is only leadership that can take us all the way.

M. Fullan (2003, p. 16)

Current realities with regard to teacher and administrator accountability in general and high-stakes testing in particular are so important in the culture of today's schools that we have added a special introduction to this subject in the second edition of *Staying on Track: An Educational Leader's Guide to Preventing Derailment and Ensuring Personal and Organizational Success*. This chapter will provide you, the reader, with a skeletal outline of the accountability/high-stakes testing situation in which assistant principals, principals, central office leaders, and superintendents find themselves. The material in this introduction provides vital information for getting the most benefit from the chapters that follow. Chapter 2 has self-assessment checklists that call for your responses to particular behaviors that often lead to derailment. Chapter 3 describes the causes of educational leader derailment, and Chapter 4 presents antiderailment strategies. Chapter 5 helps you construct personal and organizational plans for improvement.

Although our main focus in the present chapter will be on administrator accountability and high-stakes testing, the effect of this pressure on teachers is inextricably related to school and school-system leadership as well as to derailment issues. A major resource for this chapter is the seminal research and writing of R. Murray Thomas in his recently published book, aptly titled *High-Stakes Testing: Coping with Collateral Damage* (2005). By *collateral*, Thomas means attendant or parallel damage. His book has a special kind of credibility, as he is not an antitestng advocate but has instead considerable experience and expertise in assessment practices and test construction. His concern "is not about testing itself but, rather . . . about badly constructed tests, the improper administration of tests, harmful uses of test results, unrealistic standards of performance, and a lack of attention to evaluation methods other than tests" (Thomas, 2005, p. 10).

Advocates of accountability measures in general and high-stakes testing in particular argued that the following benefits would emerge. The achievement gap between the “haves” and the “have-nots”—primarily children of color and children from low-income homes—would be narrowed. Expectations for the “have-nots” would be realizable, thus stimulating teachers to do their best to reach such children. Traditional generalizations about students from good homes being successful and children from low-income homes being unsuccessful would be challenged rather than simply accepted. Teaching communities would be constructed to reach important goals with collaborative teaching and learning serving as an important vehicle for raising student achievement. Scores would go up as students and faculty discovered they could overcome what were previously felt to be obstacles. An argument for the No Child Left Behind Act (NCLB) of 2001 was that it would force many districts, principals, teachers, students, and communities to examine their areas of weakness in regard to addressing the educational needs of all students. As might be expected, however, research findings on such matters are mixed, with critics focused on collateral damage.

Professor Thomas (2005) cites four levels where administrators must apply their coping strategies: (1) national, (2) state, (3) district, and (4) the individual school.

THE NATIONAL LEVEL

Pressure at the national level is felt primarily from the alliance of the political party in power and the U.S. Department of Education as evidenced in NCLB. On January 8, 2002, President Bush signed NCLB into law. This act is the most significant reform of the Elementary and Secondary Education Act (ESEA) since it was enacted in 1965. The major stated purpose of NCLB is to narrow the achievement gap between middle-class white students and disadvantaged and minority students. The four basic emphases of NCLB are (1) stronger accountability for results, (2) more flexibility and local control, (3) greater options for parents, and (4) best teaching practices.

4 STAYING ON TRACK

The two main kinds of collateral damage that evolved are “diminished public faith in the *No-Child* plan and resentment at federal intrusion into states’ rights” (Thomas, 2005, p. 148). Meddling and inadequate funding eroded the credibility of those national politicians who mandated the program.

A major example of meddling that received tremendous media attention in January of 2005 was the disclosure that the Bush administration had paid Armstrong Williams, a prominent black media commentator, \$240,000 to plug its education policies to minority audiences. Williams was expected to produce ads that featured Education Secretary Rod Paige and that promoted Bush’s NCLB law. Williams apologized for his mistake in judgment, but said that he did not break federal law that bans the use of public money on propaganda. His critics disagreed.

Although polls of voters show enthusiasm for testing, “tests—and the sanctions that kick in when too many students fail them—are unpopular with Republican conservatives, who see *No Child Left Behind* as an unnecessary federal intrusion” (Kronholz, 2005, p. D5). Democrats who supported NCLB criticized President Bush for inadequate funding. State legislators caught “political flak when thousands of their schools didn’t meet federal achievement targets” (Kronholz, 2005, p. D5).

One of the best critiques of NCLB was written by Norman Mailer in *Parade* magazine on January 23, 2005:

NCLB would call for an unholy emphasis on doing well in tests. This could produce a narrowing of educational goals. Answers to true-or-false or multiple-choice questions would become the drill and the ability to write essays might fall to the side. That was bound to aggravate another weakness: High school students were showing reduced interest in books. (p. 5)

Mailer goes on to write about the importance of concentration in an activity such as reading. The ability to read depends on the desire to read. Furthermore, the constant interruptions in commercials in television programs every few minutes erode students’

ability to concentrate and encourage obesity as children head to the kitchen for snacks.

In an effort to stem the tide of criticism of federal involvement in high-stakes testing legislation, the U.S. Department of Education sent members of their staff on appeasement missions. Specific criticisms remained with “states and local districts’ complaints that the federal government has (a) underfunded obligatory programs, (b) set unreasonable test-performance standards for disabled and limited-English pupils, and (c) imposed unreasonable annual test-score targets, particularly for schools that enroll large numbers of children from economically disadvantaged homes” (Thomas, 2005, p. 149). A major criticism of the federal government’s role in high-stakes testing is that it has mandated a one-size-fits-all approach to curriculum formation and evaluation methods. In fact, “when accountability and standards were first introduced without much knowledge of how best to implement standards . . . , leaders accomplished little other than alienating the better teachers with unhelpful intrusions” (Fullan, 2003, p. 6).

The role of teachers unions in relation to federal legislation in general and to NCLB in particular is especially interesting. The unions are opposed to NCLB, arguing that this arbitrary and capricious measurement of teacher performance is simply wrong-headed. “No other interest group can match [the teacher unions’] political arsenal. It is not surprising, then, that politicians at all levels of government are acutely sensitive to what the teachers unions want” (Moe, 2005, p. A12). At the end of February 2004, then-Education Secretary Rod Paige called the 2.7-million-member National Education Association a “terrorist organization” after the union criticized the implementation of NCLB (Thomas, 2005). His comment was made at a meeting for the nation’s governors at the White House. Afterwards, Secretary Paige apologized for using inappropriate words.

It is clear that the education secretary has a good deal of power to set the tone in relating to various constituencies and to use discretionary measures. For example, Education Secretary Margaret Spellings, who succeeded Secretary Paige, has been

6 STAYING ON TRACK

willing to work with state and local officials in some ways that are quite different from her predecessor. She is committed to balancing states' rights to control schools with the federal government's responsibility to reduce the achievement gap between suburban white and urban minority students. She has especially listened to criticisms of NCLB by Republican politicians.

Terry Moe, a senior fellow at the Hoover Institution and a professor of political science at Stanford University, argues that the unions simply want to further their own interests: more spending, higher salaries, smaller classes, more professional development, and so on. "There is no evidence that any of these is an important determinant of student learning" (Moe, 2005, p. A12). Union contracts, he argues "are filled with provisions for higher wages, fantastic health benefits and retirement packages, generous time off, total job security, teacher transfer and assignment rights, restrictions on how teachers can be evaluated, restrictions on non-classroom duties, and countless other rules *that shackle the discretion of administrators* [italics added]" (Moe, 2005, p. A12).

An anonymous reviewer of the second edition to *Staying on Track* argues that Moe's attack on teacher unions doesn't tell the whole story: "While much of what Moe says is true, administrators also benefit from union-negotiated contracts. Administrators have higher wages, fantastic health benefits and retirement packages because of union activities. Administrators also have the ability to be more flexible with their time than teachers since they put in so many hours."

THE STATE LEVEL

Each state, according to NCLB, is expected to create standards for what a child should know and learn in reading and math in Grades 3 through 8. These standards are expected to drive the curriculum. Once these standards are set, student progress and achievement are expected to be measured annually according to state tests in a way that is consistent with state standards. Test data on each school should be publicly announced in an annual

report card. Tests are to be designed to give each teacher, school administrator, and parents data about each student's progress. Policy makers have access to these data to assess student and school success and failure. Under the provisions of NCLB, each state has the responsibility to determine what students should learn in each grade. Students who fall behind are expected to have access to special resources such as tutoring and summer school.

Administrators at the state level have been heavily involved in high-stakes testing to the extent that some educators believe that state departments of public instruction have simply become testing centers. Their coping strategies have, according to Thomas (2005), included the following:

(a) furnishing extra help to low-scoring schools, (b) providing fix-up teams for failing schools, (c) identifying improvement factors, (d) advocating a growth model, (e) preferring states' own evaluation systems, (f) reanalyzing school-performance data, (g) adopting alternative tests for special students, (h) urging Congress to abolish the testing law, (i) delaying the increase in standards, (j) fudging cutting scores, (k) not reporting failing schools, (l) granting waivers, and (m) altering teacher-qualification rules. (p. 151)

The question remains, "How will the U.S. Department of Education determine whether states have met NCLB standards and assessment requirements?" A peer review process involving experts in the field of standards and assessments will evaluate state assessment systems against NCLB requirements. In short, the peer review process will be used to examine characteristics of a state's assessment system in relation to NCLB requirements. Direct examination of a state's academic standards, assessment instruments, and test items will not take place. Instead, the peer review process will examine evidence compiled and submitted by each state to demonstrate that its assessment system meets NCLB requirements. A state may use criterion-referenced assessments and assessments that yield national norms in its academic assessment system if they are used in accordance with Department of

8 STAYING ON TRACK

Education guidelines. The peer review team will write a consensus report based on its examination of the evidence submitted by the state. The role of the state in organizing this evidence is obviously an important one.

Federal government officials in the U.S. Department of Education are often viewed by officials in state Departments of Public Instruction as outsiders determined to impose their will on the bureaucratic rung below them. School district administrators, in turn, often view officials from the state Departments of Public Instruction in a similar manner. In the process, the quality of the fix-up bureaucrats is often questioned by those below them. When large numbers of students fail tests, “students become discouraged, more drop out of school, and schools suffer punitive sanctions” (Thomas, 2005, p. 152).

THE SCHOOL-SYSTEM LEVEL

School-system administrators feel they are the third rung in the high-stakes testing bureaucracy. Many school systems had in place sophisticated assessment programs before NCLB was mandated. As a result, school-system administrators felt the new evaluation system imposed on them was a step down. Confusion was added to this matter when school-system administrators felt they were getting contradictory mandates from the federal and state governments. “If the results of federal and state testing standards are reported separately, parents can be confused about which report is the more valid whenever there is a discrepancy between the two” (Thomas, 2005, p. 154). In some cases, “schools are judged unsatisfactory under the federal plan but not under a state program” (Thomas, 2005, p. 154). It is then left to school-system administrators to try to reconcile these discrepancies.

Local assessments or a combination of state and local assessments can be used if they demonstrate that their system has a rational and coherent design. What does this mean? It means that the school system must identify the assessments, indicate how this assessment plan is aligned with the state’s academic content standards, and demonstrate how information regarding student

progress is related to the state's academic standards. The state must be able to defend the local assessments system when the peer review team visits the school system and determines if the local assessments system is consistent with NCLB guidelines.

Many school-system administrators feel that too much emphasis on reading and math has crowded out subjects such as the arts, social studies, and languages. In one district, other problems emerged as district administrators tried to align district standards with state standards. The state had set broad learning goals and had constructed tests to judge how well students were meeting them. The district then created a curriculum and a series of classroom tests to be sure that students were moving toward the achievement of state goals. Critics of these alignment procedures argued that the curriculum was shrunk to fit itself. In other words, only what was measured got done. Teachers taught to the test, thus limiting learning to only those items relevant to it. District tests were formatted to match state tests. Playing the testing game was especially pronounced on the part of beginning teachers, who were anxious about surviving in a highly competitive environment. Some teachers left teaching because of externally imposed standardization of curriculum and instruction. They felt that their professional judgment was replaced by politicians' mandates as translated by testing people in the state department of education (Brubaker, 2004).

Another problem, according to critics of the new mandates, is that high-stakes testing brings out the worst in competitive educators, some of whom turn to one or more "testing irregularities"—a euphemism for cheating. In some cases, a school's overall scores are raised by excluding certain children. A central office testing coordinator and a few teachers in one system were forced into retirement and an assistant principal was fired because of such alleged behavior.

THE SCHOOL LEVEL

School administrators, teacher leaders, and teachers are in the organizational culture where high-stakes testing takes place. They

10 STAYING ON TRACK

have to respond to critics' charges that standardized tests are too hard or too soft. NCLB requires a percentage of students at each school to score at grade level on state tests. Principals, or those who assist them, must disaggregate students by race, socioeconomic status, handicap or ability, and so forth to determine if subgroups are meeting requirements. Schools that repeatedly do not meet these goals must offer transfers to other schools or additional tutoring. With time, some teachers and administrators can be replaced and the school restructured. In North Carolina, a pioneer in the testing movement, critics say the statewide testing system is too soft. The standards were initiated 12 years ago. In the 1980s, the state legislature revamped its standard curriculum, and the ABC's accountability program was introduced in the 1990s to see how well the schools taught curriculum materials and how much students learned. What is meant by grade level is simply not clear, critics say.

Teachers and their administrators, particularly in low-income areas, often complain that their schools "focus narrowly on 'basic' academic skills, testing and discipline. The student boredom and academic failure that follow prompt calls for yet more testing and discipline" (Rabkin & Redmond, 2005, p. A9). It is also said that excessive attention to reading and math crowds out the arts. It is certainly true that at the heart of NCLB are the requirements that each state develop academic content and student achievement standards in reading/language arts and mathematics and an aligned assessment system that measures student achievement toward meeting those standards in each of Grades 3 through 8 and once in Grades 10 through 12 by the 2005-2006 school year. Some educational leaders have met this challenge by finding ways to integrate the arts into the basic academic program. "A study of 23 arts-integrated Chicago schools showed test scores rising up to two times faster than in demographically comparable schools. A study of a Minneapolis program showed that arts integration has substantial effects for all students with the greatest impact on disadvantaged learners" (Rabkin & Redmond, 2005, p. A9). It is also claimed that student progress and learning extend beyond test scores and basic subjects, so that students acquire higher-order thinking skills and feel more motivated to learn.

Michael Fullan (2003) points to a major problem that has emerged in the accountability and high-stakes testing movement:

In the 1990s, when some systems (still the minority) began using better knowledge and investing in capacity-building training of principals and teachers, there were some basic improvements, for example, in literacy and mathematics. But because these strategies were tightly orchestrated from the center, principal and teacher ownership—the kind of ownership that would be necessary to go deeper on a sustained basis—did not exist. (p. 7)

CONCLUSION

The purpose of this chapter has been to provide a skeletal outline of accountability and high-stakes testing pressures on school-system and school leaders. These pressures obviously influence the leaders' ability to stay on track and avoid derailment. You, the reader, will see in the following chapters how these matters play out in the professional lives of assistant principals, principals, central office leaders, and superintendents. School and school-system leaders who are aware of the key issues surrounding accountability and high-states testing and who have a well-reasoned plan for dealing with these issues will enhance their chances of staying on track and avoiding derailment. Leaders who do not will be a performance risk to themselves and the systems of which they are a part. Our challenge throughout this book is to find ways to create a culture of caring and excellence while at the same time dealing with matters of accountability and high-stakes testing.

Note: See Resources A through E at the end of this book for a discussion of issues related to the challenge mentioned at the end of this chapter.