

# Why Leadership Evaluation Is Broken **1**

## Topics Presented in Chapter 1 Include

- ✓ The Knowing-Doing Gap
- ✓ Progress in Leadership Evaluation
- ✓ Ambiguous Leadership Standards
- ✓ Incoherent Leadership Evaluations
- ✓ Authority–Responsibility Disequilibrium
- ✓ A Better Way: Multidimensional Leadership Assessment

Changing leadership evaluation is going to be only slightly less difficult than those most insurmountable school challenges of revising the schedule and altering the grading system. Because schools persist in the use of schedules and grading systems with which they are most comfortable (that is, which most resemble the grading systems and schedules of previous eras), the prognosis for effecting change in leadership evaluation is grim. Why then do I persist in tilting at these windmills? Because leadership evaluation at present is a mess. In the course of researching this book, I reviewed hundreds of leadership evaluation systems and descriptions of leadership evaluation procedures from active educational leaders. In general, I found prevailing leadership evaluation systems to be the “perfect storm” of failure. In his book by that title, Sebastian Junger (1998) defined the perfect storm as one in which many different variables come together at the same time to create particularly destructive consequences. The perfect storm in leadership evaluation is in evidence when there is a combination of a national leadership crisis occasioned by an acute and growing shortage of educational leaders, accompanied by a leadership evaluation system that simultaneously discourages effective leaders, fails to sanction

## 2 ASSESSING EDUCATIONAL LEADERS

ineffective leaders, and rarely considers as its purpose the improvement of leadership performance. I studied thousands of pages of documents in search of an example worthy of emulation. These leadership evaluation systems do not come from the pens of incompetent bureaucrats. These are intelligent and thoughtful people. In many cases they are my friends. But in almost every case, the evaluation systems they use are deeply flawed. These systems fail to recognize excellence, give encouragement to bad practice, tolerate mediocrity, turn a blind eye to abusive practice, accept incompetence, and systematically demoralize courageous and committed leaders. The examples that follow will clearly show that these descriptions are not hyperbole, and my willingness to challenge the present form of leadership evaluation represents my confidence that friends and colleagues would rather forgive my candor than condone a continuation of leadership evaluation as we know it.

How bad is it really? Some educational research is equivocal in tone and circumspect in conclusion. This will not be such a book. The plain truth is that educational leadership evaluation is a failure in the vast majority of cases we studied. More than 18 percent of leaders we studied had never received an evaluation in their current position. In the words of one of our research subjects, "The worst evaluation experience was no evaluation at all. The message was that I was not important enough for my supervisor to take time to give me an evaluation." What of the 82 percent of leaders who did receive at least some evaluation? The vast majority of respondents found leadership evaluation to be inconsistent, ambiguous, and counterproductive. Thus, although we know that feedback is one of the most powerful mechanisms to influence performance (Hersey, Blanchard, & Johnson, 2000; Marzano & Pickering, 2001), educational organizations use this powerful tool badly or not at all.

The research at the heart of this book, the National Leadership Evaluation Study, was conducted in 2002 by the Center for Performance Assessment. The study was based on interviews, surveys, and documentary reviews. More than 500 leaders from 21 states were included in the survey, and more than 300 leadership evaluation instruments were reviewed. Respondents had the opportunity to elaborate on their answers, and the combination of objective and extended responses forms data on which the conclusions of the study are based. In collecting leadership evaluation instruments, there was no desire to find egregious examples of poor evaluation practice. Schools shared with us the best evaluations that they had, and many of these evaluations were established as the result of thoughtful consideration and, in the case of very large systems, collective bargaining between administrator associations and the district. Despite our best efforts, this (and any study) has some limitations that should be acknowledged from the outset. The study certainly is not comprehensive and does not represent every school system in the nation. There are, to be sure, examples of excellent leadership evaluation instruments. Since the study was completed, two notable school systems, Virginia Beach, Virginia, and Omaha, Nebraska, have shared with me particularly strong leadership evaluation policies. Nevertheless, the preponderance of the evidence suggests that, at the very least, school systems should critically examine their own leadership evaluation

instruments and compare them to the best practices in assessment. Where standards are ambiguous, feedback is late, and evaluation is destructive, no national study is required to suggest that change is necessary. The details of the National Leadership Survey, including the narrative comments of the respondents, are in Resource B.

## THE KNOWING-DOING GAP

One recent survey of educational leaders revealed the growing chasm between what we know to be important and how leaders actually behave. This gap provides clear evidence that the evaluation systems that are now in place display an intellectual understanding of what needs to be done, but lack the fundamental ability to act on that knowledge. For example, the North Carolina Center for School Leadership Development (2001) found that while 60 percent of leaders strongly agreed that leaders must present evidence that their leadership vision is shared, only 30 percent of respondents performed this function. More than two-thirds of the leaders thought it important to manage time to be an instructional leader, but only 28 percent actually did so. Three-fourths of them knew it was important to collect data to develop instructional strategies and improve the effectiveness of classroom instruction, but only 40 percent strongly agreed that they performed such a function. Sixty-eight percent of the respondents strongly agreed that leaders must use the vision to guide and define decisions, but only 31 percent expressed a similar level of agreement that they performed this function. Seventy percent agree that leaders should “maintain a steady flow of two-way communications to keep the vision alive and important” and only 27 percent devoted time to such an important objective. The disconnection between expectation and reality was brilliantly captured by Jeffrey Pfeffer and Robert Sutton (2000) in their landmark book in which the title matched the subject: *The Knowing-Doing Gap*. The remainder of the survey pointed to enormous gaps between what leaders know and what they actually do. The only instance in which the gap was closed was, of course, the area that consumes the attention of most leaders—procedures and discipline. Seventy-five percent of leaders know that they should “develop and distribute student handbooks with information about rules, requirements, and expectations for student conduct and potential consequences” and 73 percent did the deed. Instructional leadership, indeed. The results of the North Carolina survey are summarized in Resource D.

Lest readers be too harsh on the state of North Carolina, it is worthy of note that the leaders in the state department of education have the courage and integrity to report these findings and work toward the creation of an improved leadership evaluation system. Many other states continue in a pattern of leadership analysis and evaluation in which they ignore glaring deficiencies. The responses by our national sample of leaders were strikingly consistent with the North Carolina findings, with the worst ratings related to the specificity of the evaluation and the relevance of the evaluation to improving student achievement.

## PROGRESS IN LEADERSHIP EVALUATION

Despite the generally deplorable state of evaluation in educational leadership, there have been some notable efforts in the right direction. The Interstate School Leaders Licensure Consortium (1996) articulated standards that have now been accepted by the majority of states, and several states have created ambitious leadership evaluation systems. Almost all of these represent an advance over previous evaluation systems and are certainly better than the utter absence of systematic leadership evaluation. It is also fair to note that the criticisms leveled at some of these systems could only take place because the authors and sponsors had the courage to publicize their systems and expose them to reviews by outsiders. Those whose leadership evaluation systems are secret or nonexistent are safe from direct criticism in this book, but their risk is far greater than that undertaken by their colleagues who had the commitment, vulnerability, and integrity to put their work on the Web and allow researchers like me to comment on it. I offer the comments in the following paragraphs in the spirit of constructive improvement and with the comment that our harshest criticism should be reserved for leadership evaluation that is shrouded in mystery, politics, and guesswork.

## AMBIGUOUS LEADERSHIP STANDARDS

The problem starts with what we call leadership, particularly in the context of education. At best, the expectations are ambiguous. At worst, the expectations are contradictory, impossible, or at great variance to common values and mountains of research. The primary problems are poorly defined standards of leadership and undefined standards of performance. There are four separate issues that plague our definition of effective leadership. First, there are poorly defined standards of leadership in which ambiguity, typically confounded by educational jargon, replaces clear expression. The second problem is undefined standards of performance, a problem that prevails even in those cases where the evaluation system has purged itself of offending jargon and ambiguity. Even the most crystal clear standard is impotent if the evaluation system does not provide a continuum of evaluation so that the adequate performance is clearly and consistently differentiated from the performance that is making progress and the performance that is exemplary. The third problem, the responsibility–authority disequilibrium, is familiar to most leaders. They are responsible for the actions of others, ranging from the most recalcitrant employee to the most disinterested community member, yet they have the authority to compel the actions of neither of these stakeholders.

In the vast majority of the leadership evaluation documents I reviewed, one of two problems prevailed. Either the standards themselves were ambiguous or the performance expectations were unclear. The following statements have been gleaned from local, state, and national expectations for school leaders. After each statement is a challenge that any leader being evaluated by such a standard would want to consider.

<b>Expectation</b>	<b>Challenge</b>
“The administrator facilitates processes and engages in activities ensuring that curricular, co-curricular, and extra-curricular programs are designed, implemented, evaluated, and refined.”	What in the world does this mean? How would we know if this standard has been met? Does evaluation and refinement have to do with what is popular or what is effective?
“Stays current with research and theory regarding theory and motivation. Keeps abreast of the latest developments in the field of education.”	<i>Any</i> research and theory? Much of it is awful and contradictory to the goals and values of the school system. This goal appears to endorse a collection of fads, the proverbial “flavor of the month” in which school leaders fail to distinguish what is current from what is important, valid, tested, trustworthy. When someone announces with breathless enthusiasm that he has found the True Path in a recent discovery, such as interdisciplinary instruction, I am compelled to ask, “Didn’t we call that ‘humanities’ 30 years ago?” When I listen outside the doors of national conventions or local staff development meetings, I hear ideas—including some good ones—that are of the vintage of Socrates or Dewey, yet are promoted as if they are copyrighted by a 21st century guru with exceptional insight. The point is not the elevation of one era over another, but rather the application of research and thinking of millennia with judicious caution, appropriate skepticism, and historical context.
“Provides information on curriculum/instruction.”	The issue is not whether the leader provides information, but rather whether the information is of sufficient quality to be understood and whether the information is used to make good decisions to improve student achievement.
“Expects and coaches effective classroom practices and a service orientation.”	I have an idea of what effective classroom practices are, but unless they are specified, the definition of effectiveness can vary wildly from one administrator to the next. The variations are as likely to be based on opinion as on research. I do not know what a service orientation means in this context.
“The administrator facilitates processes and engages in activities ensuring that relevant demographic data pertaining to students and their families are used in developing the school mission and goals. Diversity is considered in developing learning experiences.”	Does this mean that good leaders have different goals for poor schools than for rich schools? Is it a good idea to have different goals for schools based on their ethnic composition? Does this mean that if the families have a culture of low expectations that schools should mirror those expectations?
“Demonstrates effective organizational skills.”	What skills? What does effectiveness mean in this context?
“Participates in professional development activities.”	My fourth grader’s hamster can participate in professional development activities. What does this tell us about the impact of using new knowledge and skills to become a more effective leader?

(Continued)

## 6 ASSESSING EDUCATIONAL LEADERS

(Continued)

<b><i>Expectation</i></b>	<b><i>Challenge</i></b>
In the same leadership evaluation instrument, one district requires that its effective leader “carefully weighs consequences of contemplated action” and, a few sentences later, requires the same leader to be “action-oriented; presses for immediate results. Decisive; doesn’t procrastinate on decisions.”	As if this were not sufficiently problematic, the rating scale for the 21st century Solomon who is to occupy a leadership position in this district has only two possible ratings: “Meets Standard” and “Does Not Meet Standards.” It is no surprise that the same evaluation form requires the leader to simultaneously “hold to personal opinions,” “exhibit a need to control most situations,” and “demonstrate adaptability and flexibility.” One might gently suggest that an administrative certificate and a doctorate are not the criteria sought by this district, but rather some combination of divinity and multiple personality disorder.

### INCOHERENT LEADERSHIP EVALUATIONS

Not every leadership evaluation instrument is so deeply flawed with regard to the establishment of clear leadership standards. But even where I found clarity in the standard, I discovered ambiguity with regard to the description of performance levels. Typical of the language of performance evaluation was “exceeds expectations” or “meets expected performance levels” or “superior” or “average”—without any clear indication of which specific leadership behaviors deserved such labels. In the absence of specification, the only criterion available is the idiosyncratic judgment of the evaluator. However wise and insightful an individual evaluator may be, these judgments are doomed to be inconsistent and of little use for coaching. The recipient of the evaluation only knows that one evaluator regarded her as “outstanding,” but the same leadership traits and behaviors merited a rating of “meets standards” from another evaluator while yet a third evaluator looked at the same performance and said that it “exceeds expectations.” Out of such linguistic mire one should not expect leadership wisdom to emerge.

Effective evaluation systems allow both the evaluator and the one being evaluated to understand clearly the difference between various levels of performance. Michael Jordan, for example, is acutely aware of the difference between putting the ball in the basket and hitting the rim. His fans share his perceptions of clarity in evaluation. Sarah Chang, along with the vast majority of her audience, knows the difference between an F-natural and an F-sharp. But do school leaders, to whom we entrust our children and billions of dollars in resources, know the difference between performance that is exemplary and that which is deplorable? Descriptions of performance—even if the standards themselves are clear—devolve into the linguistic quicksand of “sometimes” compared to “seldom” or “frequently” compared to “often” or “exceeds expectations” compared to “satisfactory.” Intelligent people of good will can disagree about what any of these descriptions mean. Perhaps the least helpful are the descriptions such as

“growth needed”—invariably a negative comment in the context of evaluation, yet I strain to think of a single leader, from Alexander the Great to Napoleon to Churchill to the very best school leaders I have observed in more than a million miles of travel who would not enthusiastically check a box called “growth needed” when describing themselves. To put it bluntly, when is growth not needed? Presumably when one is dead.

It’s really not as bad as all that—it’s worse. Consider the following examples of descriptions of performance:

<b><i>Expectation</i></b>	<b><i>Challenge</i></b>
“The principal always meets and sometimes exceeds expectations for performances in this position requirement.”	Whose expectations? How frequently? What does “sometimes” mean? What is the difference between “meeting” and “exceeding” expectations? If the expectation is fair and ethical behavior, how does one exceed it? If the expectation is that the principal provides equitable opportunities for 100 percent of students, how is it possible to exceed such a standard?
“Students using technology or products of technology. Yes or No.”	By such a binary standard, students gain the same credit for an insightful analysis of literary criticism of Hemingway discovered on the Internet as they do for playing Nintendo or e-mailing a friend.
“No disruptions. Yes or No.”	Are these people serious? When is the answer “yes?” In the summer? At midnight? In the cemetery? In a school without students?

I could go on, but the point has been made. Schools are succeeding in spite of, rather than because of, these tainted leadership evaluation instruments. If they could continue to depend upon a generation of leaders who would persist, learn, and succeed without any meaningful coaching, reinforcement, or sanctions from their evaluation system, then this book would be unnecessary. But in a nation in which a growing number of leadership positions are unfilled and an alarming number of leaders are leaving the field of education, it is well past time for fundamental reform in a system that is not merely broken, but shattered. The nation needs a new form of leadership evaluation and it needs it now.

## **AUTHORITY–RESPONSIBILITY DISEQUILIBRIUM**

We wish our leaders to be some mythical combination of folk heroes, in which they have the insight of Lao-Tzu, the courage of a New York firefighter, the risk tolerance of Amelia Earhart, and the work ethic of Paul Bunyan. For the purposes of rhetorical flair and savvy marketing, publishers cast as wide a net as possible, offering us models for leadership who include, in alphabetical order, Attila the Hun, Catherine the Great, Churchill, Elizabeth I, Jefferson, Jesus, Machiavelli, Moses, Napoleon, Nixon, Rasputin, Roosevelt (Teddy and Franklin), Washington, and untold numbers of yet to be reconstructed historical leaders whose biographers have found some link between personal traits

## 8 ASSESSING EDUCATIONAL LEADERS

and organizational effectiveness. In the real world of school leaders, however, the relationship between demands and authority leads to more prosaic results. This does not stop the authors of leadership evaluations from the grand presumption that the school principal or district superintendent enjoys monarchical powers.

The most glaring examples of the authority–responsibility disequilibrium occur when the administrator is held responsible for the actions of others. One set of leadership standards, for example, requires that the leader “ensures that staff and community understand the analysis of student data.” The leader can provide information to the staff and even assess the staff’s knowledge, but cannot “ensure” understanding. Certainly the leader can share information with the community, but that is a long way from guaranteeing comprehension or even interest by the general public. Another leadership evaluation requires leaders to “ensure a balanced budget.” This might require controlling the weather in New England, the price of oil in Iraq, the impact of hail on the roofs of schools in Kansas, and the number of snow days in Idaho. A host of natural events can affect the budget in ways that are far beyond the control of school leaders.

### **A BETTER WAY: MULTIDIMENSIONAL LEADERSHIP ASSESSMENT**

There is an alternative to the vacuous exercises now called leadership evaluation. Building on the best practices in performance assessment, this book offers a new model of leadership evaluation: multidimensional leadership assessment (MLA). In assessing any human performance, whether a student leaning to read or a superintendent leading a complex system, there is a complex set of variables that must be assessed. Imploring students to read well is of little more value than exhorting the superintendent to be a better leader. Improvement requires feedback that is specific, accurate, and timely. Thus, rather than making evaluation an event that occurs once every year (or in the case of senior leaders, every three or four years, always after it is too late to influence performance), MLA provides frequent feedback with multiple opportunities for continuous improvement. Rather than providing performance levels that are meaningless (“meets expectations,” “above average,” or “progressing toward standards”), MLA describes in specific terms the difference between performance that is distinguished and performance that is proficient, progressing, or failing to meet standards.

Finally, MLA is not merely a retrospective approach to leadership evaluation. MLA can be used to improve the performance of a 30-year veteran and to coach the newest assistant principal. It can also be used to train new leaders and identify prospective leaders who will become the next generation of educational leadership. Most important, MLA will force school systems to establish clear, coherent, and fair expectations for present and future leaders. In the past three decades, tremendous strides have been made in every area of educational assessment. It is well past time that we apply those lessons to the assessment of educational leaders.