

SECOND EDITION

ASSESSING EDUCATIONAL LEADERS

Evaluating Performance for Improved
Individual and Organizational Results

DOUGLAS B. REEVES



Contents

Preface to the Second Edition	ix
Acknowledgments	xiii
About the Author	xvii
1. Why Leadership Evaluation Is Broken	1
The Knowing-Doing Gap	3
Progress in Leadership Evaluation	4
Ambiguous Leadership Standards	4
Incoherent Leadership Evaluations	6
Authority–Responsibility Disequilibrium	7
A Better Way: Multidimensional Leadership Assessment	8
2. Reframing Leadership Evaluation	9
Patrick’s Choice: The Risks and Rewards of Evaluation	9
A New Vision of Educational Accountability	13
Criteria for Multidimensional Leadership Assessment (MLA)	16
Implications of a New Vision for Leadership Evaluation	21
3. Moving Beyond One-Dimensional Leadership Evaluation	25
The Prevalence of One-Dimensional Evaluation	25
Leadership Is More Than Test Scores	27
Leadership Is More Than Popularity	29
Leadership Is More Than Obedience	32
4. Creating an Improved Leadership Evaluation System	35
Domains of Leadership Performance	37
Leadership Performance Standards	43
Multidimensional Leadership Assessment	46
5. Using Evaluation to Improve Performance	55
The Capacity-Building Continuum	56
The Challenge Continuum	60
The Coaching Continuum	61
The Counseling Continuum	61
Defining Performance: The Key to Using the Leadership Evaluation Continua	62

6. Developing a Multidimensional Leadership Assessment System	65
Leadership Domains	66
Leadership Performance	67
Leadership Assessment as a Research Goldmine	69
Organizational Goals, Values, and Priorities	70
The Performance Continuum	74
Qualitative Information: The Lens for Understanding	
Quantitative Data	75
Reflection and Self-Regulation	76
7. Building the Next Generation of Educational Leaders	79
The Leadership Development Imperative	79
Prototype for a Leadership Development System	80
Authentic Assessment for Future Leaders	81
8. Leading Leaders	87
The Coaching Conundrum: Results or Antecedents?	87
Inquiry and Hypothesis Testing	89
Applying MLA Throughout the Organization	91
9. Improving Leadership Evaluation With Multidimensional Leadership Assessment	93
Analyze the Present Evaluation System	94
Create Dimensions of Leadership	95
Specify Performance Levels	96
Field Test the Instrument at Multiple Levels	97
Implement Throughout the Organization	97
Share Results With Universities and Research Institutions	98
Use MLA for Training, Recruiting, and Developing Leaders	99
Conclusion: Finding Answers Inside	100
10. Senior Leadership Assessment: The Hallmarks of Excellence Leadership Research	101
Challenges of Senior Leadership Assessment	101
Assessment <i>for</i> Leadership: The Hallmarks of Excellence in Educational Leadership	102
The Fundamentals of Hallmarks	102
From Assessment to Improvement:	
Leadership Performance Coaching	104
Conclusion	105
11. Leadership Responsibilities for Planning, Implementation, and Monitoring	107
The PIM Research	107
How Leadership Plans Improve Achievement	108
Conclusion	111
Resource A: The Leadership Performance Matrix	113
Resource B: National Leadership Survey Results	137

Resource C: Leadership Evaluation Survey	151
Resource D: The Gap Between What Leaders Know and What They Do	155
Resource E: Principal Evaluation Rubrics	169
Resource F: Ascension Parish Leadership Professional Growth Matrix	187
References	213
Index	215

Preface to the Second Edition

In the first edition of *Assessing Educational Leaders*, I claimed that leadership evaluation was a mess. Our research revealed that the evaluations of educational leaders were frequently inconsistent, ambiguous, and unrelated to the strategic objectives of the school system—and that was when evaluations happened at all. In almost 20 percent of the cases we studied, leaders had never been evaluated in their present position. Finally, we found that the longer the tenure of leaders in their current position and the greater their responsibilities within a school system, the less likely they were to receive accurate and constructive evaluations. The response to these findings from a broad range of educational leaders, policymakers, advisers, and researchers suggested that it was time to update the book, add new resources for readers, and provide case studies of success.

In this new edition, I offer evidence that a growing number of school systems are making significant improvements in their leadership evaluation procedures, providing models for the educational world to consider. Moreover, researchers and scholars offer practical insights into the key distinction between *evaluation* of leaders—a process sometimes fraught with politics, subjectivity, and relationship-poisoning judgment—and *assessment* of leaders—a process designed to provide feedback that will improve leadership performance. That distinction is at the heart of the new content in this edition, including

- Principal Evaluation Rubrics (Resource E), an exceptionally creative contribution to the field by Kim Marshall, leadership coach for New Leaders for New Schools and the editor of *The Marshall Memo* (www.marshallmemo.com). Marshall not only brings a singular grasp of educational research to this project but provides the most practical method of principal evaluation I have found.
- Hallmarks of Excellence Leadership Research (Chapter 10), a cutting edge leadership assessment and coaching tool designed to provide confidential feedback to senior leaders.
- Planning, Implementation, and Monitoring (PIM) Research (Chapter 11), revealing the specific actions of educational leaders that are most linked to improved student achievement.

- Examples of real-world applications of the Ascension Parish Leadership Professional Growth Matrix found in Resource F.

While the new evidence in this edition offers some cause for optimism about the potential for improved leadership assessment, there is also considerable cause for caution. First, in the United States alone, we are about to witness a leadership turnover of unprecedented proportions, with the American Association of School Administrators (Davis, Darling-Hammond, LaPointe, & Meyerson, 2005) estimating that more than 40 percent of school leaders will be eligible for retirement within the next four years. Worse yet, the schools and districts most in need—poor, urban, and exceptionally challenging—are those least likely to retain effective leadership. Even among the nation's leading urban school systems, according to the Council of the Great City Schools, superintendent tenure averages only 3.1 years (2006), and cases of urban schools and districts with revolving doors in the executive suite are common.

Some of these challenges are systemic—high-need schools and districts can burn leaders to a cinder with unsustainable hours and extraordinary stress, and therefore higher turnover might come with the territory. But many of the challenges causing leadership turnover are self-inflicted wounds. In particular, boards of education place demands on superintendents and, in turn, superintendents place demands on subordinate leaders, that range from the unreasonable to the ridiculous: The superintendent reprimanded by the board for failure to attend the right service club meetings; the principal called on the carpet for attending the birth of twins rather than the right basketball game; the academic dean raked over the coals for requiring a student to participate in a reading intervention that prevented a failure but irritated an activist parent. If this book can be reduced to a single sentence, it is the following: *Leadership assessment must be focused on effectiveness, not popularity.*

Improvements in student assessment are a hallmark of the past two decades of educational research and practice. From a tradition dominated by multiple-choice tests and norm-referenced assessments, the work of Wiggins (1998) and Wiggins and McTighe (2005), Darling-Hammond (1997), Stiggins (2000), and Stiggins, Arter, Chappuis, and Chappuis (2004) have brought authentic assessment and assessment *for* learning into the mainstream. It is therefore deeply ironic that the part of the educational establishment where advanced degrees predominate—building administrators and senior leadership—has failed to keep pace in the use of assessment that is designed to improve performance. If first-year teachers provided feedback to students in a manner that was ambiguous, inconsistent, and unrelated to performance goals, then their jobs would be in jeopardy. But if elected officials and administrators with terminal degrees commit the same offenses with their evaluations, then the too common reaction is resigned acceptance. Thus this edition of *Assessing Educational Leaders* is a clarion call to action. We should expect no less of policymakers and senior administrators than we require of novice teachers—evaluation, assessment, and feedback that is accurate, specific, and clear. Most important, we must change the fundamental purpose of assessment of leaders, following

the pattern of recent changes in the assessment of students. The purpose of assessment is not to rate, rank, sort, and humiliate. The purpose of assessment is to improve performance. Only when leadership assessment achieves that goal will this book have achieved its purpose.

—*Douglas B. Reeves*
Salem, Massachusetts
January 2008

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

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.

<i>Expectation</i>	<i>Challenge</i>
“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)