# Kim Marshall

# RETHINKING TEACHER SUPERVISION AND EVALUATION

28

How to Work Smart, Build Collaboration, and Close the Achievement Gap

SECOND EDITION

### More Praise for Rethinking Teacher Supervision and Evaluation

"Kim has written an important book, a must-read for anyone who believes that school can be and must be greatly improved. His argument is clear, cogent, and buttressed by endless practical ideas borne of his long work on the front lines of urban education. But his message applies to all educators: pay attention to what does and doesn't work; build upon a rich curriculum; and, most of all, redefine supervision and evaluation to ensure that they are ongoing, efficient, and always focused on staff learning about learning. Marshall has written a thoughtful, practical, and hopeful book that should play a vital role in ending the fatalism that is a cancer in American education."

— Grant Wiggins, president, Authentic Education, Hopewell, New Jersey

"Kim Marshall is quickly establishing himself as a preeminent voice in the field of school leadership. He does what few authors have done before: integrate the critical tasks of observation, curriculum planning, and data-driven instruction to drive real student achievement. His combination of real-world experience and visionary thinking creates a road map that has the potential to alter the national landscape on teacher supervision."

— Paul Bambrick-Santoyo, managing director, North Star Academy Network, Uncommon Schools

"Classroom teachers and principals alike will decorate the pages of this insightful book with notes and underscored passages that offer inspiration and the kind of down-to-earth advice you seek from the best teacher you know. Kim Marshall is a visionary thinker with his finger right on the pulse of how good teaching transforms thinking, and how deep concern for students transforms teaching; in direct, compelling language, he describes the magical synergy that happens when educators at all levels work together to prepare students for life—not just for tests."

— Vicki Spandel, author, Creating Writers and The Nine Rights of Every Writer

"Kim Marshall brings the wisdom of a seasoned principal and the insights of a scholar-researcher to this analysis of instructional leadership. His work is practical, smart, and, most of all, clear and accessible. Any educational leader seeking to find ways to improve the quality of instruction will find in this book an invaluable resource. Marshall writes with the authority of someone who understands what needs to be done to create successful schools because he has already done so himself."

 Pedro A. Noguera, Ph.D., Peter L. Agnew Professor of Education, Steinhardt School of Culture, Education, and Development, executive director, Metropolitan Center for Urban Education, New York University

"Kim Marshall's invaluable book is more than a new view of supervision and evaluation—it offers a broad-based map of the multiple paths that a principal must consider if teaching and learning are to be improved deeply."

-Jon Saphier, author, The Skillful Teacher

"In this useful volume, Marshall draws upon his years of experience doing the work as a principal, coach, and principal-educator to offer a clear-eyed critique of both business-as-usual teacher evaluation as well as a number of faddish proposed approaches. In their place, Marshall offers administrators a pragmatic, concrete, and effective system for carrying out supervision and evaluation that can genuinely move the dial on teaching, learning, and student achievement. Far from a dull primer, this text reads like a coach at the school administrator's side. I am already drawing upon Marshall's proffered system and strategies with my own aspiring educators and their university supervisors."

— Scott Seider, assistant professor of education,
Boston University

# RETHINKING

# TEACHER SUPERVISION AND EVALUATION

H

How to Work Smart, Build Collaboration, and Close the Achievement Gap

SECOND EDITION

Kim Marshall

**IB** JOSSEY-BASS<sup>™</sup>
A Wiley Brand

Cover design: Michael Cook

Cover illustration: © Boris Hajdarevic/iStockphoto

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by Jossey-Bass A Wiley Brand One Montgomery Street, Suite 1000, San Francisco, CA 94104-4594—www.josseybass.com

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, or on the Web at www.copyright.com. Requests to the publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, 201-748-6011, fax 201-748-6008, or online at www.wiley.com/go/permissions.

Permission is given to reproduce the rubrics in this book for educational purposes. Systematic or large-scale reproduction or distribution (more than 100 copies per page, per year), electronic reproduction, or inclusion in any publications offered for sale or used for commercial purposes are all subject to the prior written permission of the publisher.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Readers should be aware that Internet Web sites offered as citations and/or sources for further information may have changed or disappeared between the time this was written and when it is read.

Jossey-Bass books and products are available through most bookstores. To contact Jossey-Bass directly call our Customer Care Department within the U.S. at 800-956-7739, outside the U.S. at 317-572-3986, or fax 317-572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

### Library of Congress Cataloging-in-Publication Data

Marshall, Kim, 1948-

Rethinking teacher supervision and evaluation: how to work smart, build collaboration, and close the achievement gap / Kim Marshall.—Second edition.

pages cm

Includes bibliographical references and index.

ISBN 978-1-118-33672-4 (pbk.)

ISBN 978-1-118-41661-7 (ebk.)

ISBN 978-1-118-41997-7 (ebk.)

ISBN 978-1-118-43415-4 (ebk.)

1. Teachers—Rating of. 2. Teachers—In-service training. I. Title.

LB1728.M26 2013

371.14'4—dc23

Printed in the United States of America

PB Printing 10 9 8 7 6 5 4 3 2

2012051513

### CONTENTS

The Author Acknowledgments	vii viii
Introduction	1
ONE The Challenge: Closing the Achievement Gap	7
rwo Supervision and Evaluation: Why We Need a New Approach	19
THREE Mini-Observations 1: A System Is Born	43
FOUR Mini-Observations 2: Doing Them Right	57
FIVE Curriculum Design: The Foundation of Good Teaching	87
SIX Interim Assessments: Using During-the-Year Evidence of Learning to Continuously Improve and Evaluate Instruction	103
SEVEN Rubrics: Potent, Efficient End-of-Year Evaluation Tools	123
EIGHT Time Management: Doing First Things First	157
NINE The Role of the Superintendent	173
ren Putting It All Together	195
APPENDIX A A Slim Curriculum Expectations Booklet: Atlantic Elementary School, Grade 4 Learning Expectations	199
APPENDIX B Examples of Brief Write-Ups after Mini-Observations	213
APPENDIX C A Comparison of Rubrics	219
Bibliography Index	221 227

# For Lillie and David, skillful and intrepid teachers

### THE AUTHOR

Kim Marshall was a teacher, central office administrator, and principal in the Boston public schools for thirty-two years. He now advises and coaches new principals, working with New Leaders (formerly New Leaders for New Schools); teaches courses and leads workshops on instructional leadership; and publishes a weekly newsletter, the *Marshall Memo*, which summarizes ideas and research from sixty-four publications (www.marshallmemo.com). Marshall has written several books and numerous articles on teaching and school leadership. He is married and has two children; both are teachers, one in Boston, the other in California.

irst and foremost, I am grateful to my wife, Rhoda Schneider, for her support, wise counsel, and keen eye, and also to Lillie Marshall, David Marshall, Katherine Marshall, and Laura Marshall.

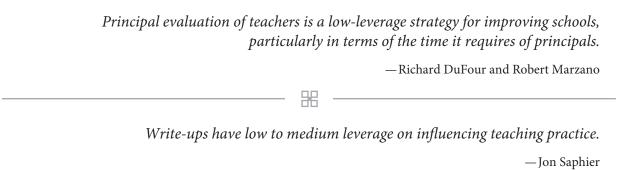
Christie Hakim at Jossey-Bass believed in this book from the beginning and persuaded me to write it, and she and her colleagues contributed mightily to the final product, including Leslie Tilley (special thanks for helping reformat the rubrics), Julia Parmer, Hilary Powers, Kate Gagnon, and Pam Berkman. The second edition has benefited from close attention from Kate Gagnon and Tracy Gallagher.

A loyal group of friends and thought partners have helped encourage and shape this book over the years: Jon Saphier, Mike Schmoker, Doug Reeves, Roland Barth, Paul Bambrick-Santoyo, Larry Cuban, Jay McTighe, Grant Wiggins, Jeff Howard, John King, Rick DuFour, Dylan Wiliam, Charlotte Danielson, Mike Lupinacci, Mark Jacobson, Jenn David-Lang, Andrew Bundy, Andy Platt, Barney Brawer, Barry Jentz, Mary Ellen Haynes, Lorraine Cecere, Dick Best, Lois Jones, Bill O'Neill, Bill Henderson, Bob Weintraub, Doug Lemov, Ellie Drago-Severson, George Hill, Gerry Degnan, Vikki Ginsberg, Jay Heubert, Joan Dabrowski, Karen Drezner, Kathleen Flannery, Mairead Nolan, Mark Roosevelt, Mary Grassa O'Neill, Mary Russo, Maureen Harris, Michael Fung, Emily Cox, Sandi Kleinman, Pamela Seigle, Pedro Noguera, Vicki Spandel, Penny Noyce, Pete Turnamian, Jamey Verilli, Rick Weissbourd, Sandy Mitchell-Woods, Ted Dooley, Diane Lande, Toni Jackson, Maria Palandra, Betsey Useem, and Mike Useem. I'm especially grateful to Paul Bambrick-Santoyo for a thoughtful critique of the new Chapter Nine in the second edition.

Finally, I am grateful to the teachers at the Mather School, who tutored me as these ideas germinated, and to the budding principals, seasoned coaches, and honchos in New Leaders for New Schools, who have contributed in ways they cannot imagine: Jon Schnur, Monique Burns, Ben Fenton, Cami Anderson, Jann Coles, Kris Klasby, Stephanie Fitzgerald, Vera Torrence, Mark Murphy, my coaching colleagues in New York; Washington, D.C.; Chicago; and the Bay Area; and all New Leaders principals.

And my special thanks to Athie Tschibelu, who went above and beyond the call of duty to help launch one of the first components of this book.

### INTRODUCTION



o many educators, these two quotes from three of America's leading authorities on instructional improvement are shocking and counterintuitive. For decades, the assumption has been that if we want to improve teaching, supervision and evaluation are effective levers. Surely, the argument went, inspecting classroom performance and giving teachers feedback and formal evaluations would make a positive difference.

But when educators take a few minutes to reflect on what DuFour, Marzano, and Saphier are saying, it begins to make sense. I frequently ask groups of administrators to think back to when they were teachers and raise their hands if an evaluation ever led them to make significant improvements in the way they taught. Typically, around 5 percent raise a hand. When I ask if the evaluations that principals themselves have written produce significant classroom improvements, I get a similar response. Most principals sheepishly admit that after all the work they put into all those pre-observation conferences, classroom visits, write-ups, and post-conferences, they rarely see much difference in what teachers do—much less in student achievement.

This is disturbing. It means that school leaders are spending huge amounts of time on a process that rarely improves classroom teaching. And teaching, after all, is the heart of the matter. Research has shown that the quality of instruction is the single most important factor in student achievement (Ferguson & Ladd, 1996; Sanders & Rivers, 1996; Sanders, Saxton, and Horn, 1997; Haycock, 1998; Rivkin, Hanuschek, & Kain, 2005; Whitehurst, 2002; Hattie, 2002; Rice, 2003; Nye, Hedges, & Konstantopoulos, 2004; Clotfelter, Ladd, & Vigdor, 2007).

### **GOOD TEACHING REALLY MATTERS**

This was not always the conventional wisdom. For years, factors outside the schoolhouse were believed to be the main determinants of how children did—social class, innate intelligence, family background, community dynamics, negative peer pressure, racism, and discrimination. But now we know that good classroom teaching can overcome the disadvantages with which many students enter school, and that children who grow up in poverty are not doomed to failure. Figure I.1 shows the dramatic difference in the achievement of students who have three years of effective, mediocre, or ineffective teaching.

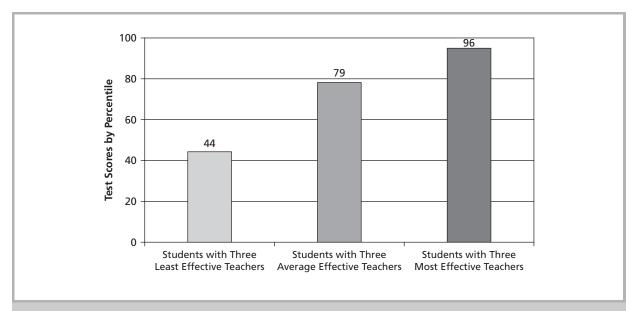


Figure I.1 Fifth-Grade Math Scores on Tennessee Statewide Test: The Difference the Quality of Teaching Makes as Similar Students Move from Grade 3 to 5 Source: Sanders and Rivers (1996).

Good teaching helps all students, but it turns out that it makes a bigger difference for some than for others. Figure I.2 shows the results of a study that compared the impact of effective and ineffective teachers on students as they moved from fifth to seventh grade. Students who were lucky enough to have effective teaching for three years in a row achieved at almost identically high levels, even though some started with much lower achievement than others. But a matched sample of students who had three years of ineffective teaching fared quite differently: those who started out with high and average achievement were still doing quite well at the end of seventh grade, but students who started out with low skills did much worse. This study and others like it show that low-achieving students benefit disproportionately from good teaching. Unfortunately, the children who need good teaching the most—those who are economically disadvantaged, members of minority groups, and those with special needs and language deficits—are more likely to attend schools with an inexperienced, transient teacher corps and unfavorable, sometimes chaotic learning environments.

So here's the logic of the preceding paragraphs: (a) teaching really matters, (b) not all teaching is equally effective, (c) teaching quality is unevenly distributed by class and race, and therefore (d) there is an inexorable, day-by-day widening of the achievement gap across the nation.

Some broad societal challenges flow from this analysis: how to get our most effective teachers teaching our neediest students; how to create working conditions that will attract them to some pretty embattled schools; and how to create an esprit in the national teacher corps similar to that among firefighters,

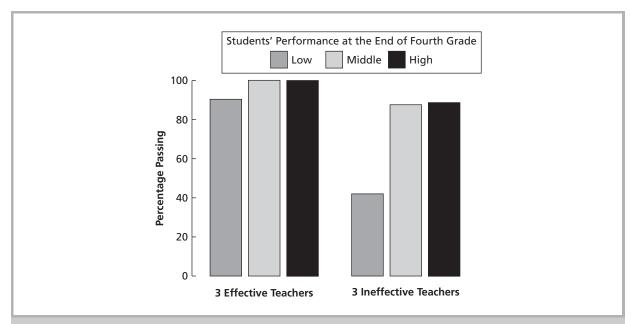


Figure I.2 The Impact of Effective and Less Effective Teachers on Grade 5–7 Students with Different Levels of Entering Proficiency

Source: Bracey (2004).

among whom the "best and bravest" want to work in firehouses with the "best" fires—that is, the most challenging ones.

But for the short term, we can draw a conclusion with which almost every parent would agree: every principal's most important job is getting good teaching in every classroom.

Which brings us to the subject matter of this book: What is the best way to get effective teaching for every child? For starters, hiring and firing. Removing *ineffective* teachers is critically important, as is hiring talented and hard-working teachers, since each vacancy is a golden opportunity to upgrade the team. But vacancies don't occur that frequently. So while hiring and firing are tremendously important, this book will focus on strategies for supporting and improving the teachers (ranging from excellent to ineffective) who are in classrooms now.

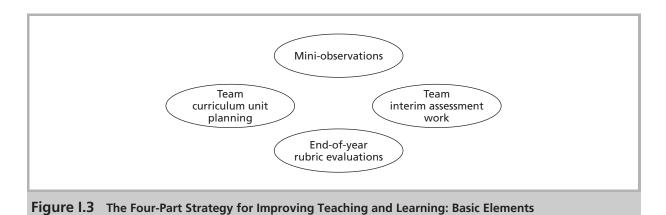
In recent years, schools and districts have tried a variety of approaches for improving teaching:

- More aggressive supervision and evaluation
- Using test scores to evaluate teachers
- Publishing teachers' test scores in newspapers
- Merit pay for high-performing teachers
- Revamping the teacher evaluation forms that principals fill out
- Doing "learning walk" or "instructional rounds" tours of schools with feedback to the staff
- Getting teachers to visit exemplary classrooms and schools
- Having teachers analyze student work
- Requiring teachers to use highly scripted curriculum programs
- Providing laptop computers for every student
- Encouraging teachers to use the Internet to find effective ideas and materials

- Setting up "critical friends groups" in which teachers read and discuss articles and books
- And the old standby, getting teachers to attend workshops and courses inside and outside their schools

Each of these approaches can contribute to the quality of instruction under the right conditions, and they all have proponents. But I believe there is a much more powerful way to improve teaching and learning and close the achievement gap.

This book will present four closely linked strategies centering on specific actions principals can take: (a) making short, unannounced classroom visits followed by one-on-one feedback conversations, (b) participating much more actively in the curriculum unit planning process, (c) working with teacher teams to analyze and follow up on interim assessment results, and (d) using rubrics for end-of-year teacher evaluation. I believe these are the most effective ways for a principal to exercise instructional leadership and make a real difference at the classroom level. Figure I.3 is a diagram that will evolve over the course of this book to show how the four strategies interact.



Implementing these four strategies involves fundamental changes in the way principals handle supervision and evaluation and the professional dynamic within schools. School leaders shift in these ways:

- From periodically evaluating teaching to continuously analyzing learning
- From infrequent announced classroom visits to frequent unannounced visits
- From taking extensive notes on one or two lessons a year to watching for key "teaching points" in each of a number of visits
- From guarded, inauthentic communication with teachers to candid give-and-take based on authentic classroom observation
- From formal yearly or twice-yearly evaluations to continuous suggestions and redirection, culminating in an end-of-year evaluation
- From inadvertently sowing envy and division among teachers to empowering and energizing teacher teams
- From teachers saying "Let me do it my way" to everyone asking "Is it working?"
- From administrators doing most of the work to teachers taking on real responsibility for improving teaching and learning
- From evaluating individual lessons to supervising the effectiveness of curriculum units

- From one-right-way evaluation criteria to constantly looking at new ideas and practices
- From focusing mainly on ineffective teachers to improving teaching in every classroom
- From cumbersome, time-consuming evaluations to streamlined rubrics
- From being mired in paperwork to continuously orchestrating schoolwide improvement

This book comes from my own experience, extensive research, and close observation of scores of effective and ineffective schools. Thirty-two years as a Boston teacher, central office administrator, and principal were the starting point. Since leaving the Boston schools in 2002, I have continued to develop my initial insights and practices and have presented them in hundreds of workshops and consultations in a wide variety of schools and districts. These ideas have been adopted in part by a number of schools.

Although there is no gold-standard research on the model, I believe the combination of all four elements is the key to dramatic gains. Mini-observations have a powerful logic, and I challenge readers to present a better system for seeing what is really going on in classrooms on a daily basis and coaching teachers throughout the year. Curriculum unit planning has a strong track record all over the world, thanks to the work of Grant Wiggins, Jay McTighe, and their colleagues. Interim assessments have robust research support and a number of successful practitioners. And teacher evaluation rubrics, while they are very recent arrivals in schools, just make sense.

Here is a chapter-by-chapter overview of the book:

- Chapter One tells the story of my fifteen-year Boston principalship, during which my colleagues and I struggled against significant obstacles and realized quite late in the game that major gains in student achievement are difficult without external standards linked to good assessments.
- **Chapter Two** describes the design flaws in the conventional supervision and evaluation process that explain why it rarely improves teaching and learning.
- Chapter Three gives the blow-by-blow of my initial failure as a principal to get supervision and evaluation working well, and my discovery of mini-observations—an effective way of getting into classrooms and giving teachers feedback.
- Chapter Four is an analysis of twelve key factors for implementing mini-observations successfully.
- Chapter Five broadens the usual definition of supervision to include principals working with teacher teams as they clarify learning goals and "backwards design" curriculum units—all of which helps teachers draw on each others' insights and wisdom and makes the principal a more perceptive and helpful thought partner during and after classroom observations.
- Chapter Six broadens supervision further still, describing how principals can direct and support teacher teams as they look at interim assessment results, figure out learning problems, help struggling students, involve students in improving their own performance, and continuously improve instruction.
- Chapter Seven returns to a major flaw in traditional year-end teacher evaluations—how time-consuming and disrespected they often are—and presents a set of teacher evaluation rubrics that are more time-efficient and give teachers succinct, detailed, and constructive end-of-year feedback.
- Chapter Eight deconstructs the time management challenge that all principals face and suggests ten ways school leaders can survive, do first things first, and get the engine of improvement humming in their schools.

- Chapter Nine suggests ways that superintendents can support and direct the work of principals as they implement this four-part model; the chapter includes a principal evaluation rubric.
- Chapter Ten sums up, describing how mini-observations, curriculum planning, interim assessments, and teacher evaluation rubrics interact and build on each other. By successfully orchestrating these four components, principals can build collaboration with and among teachers and foster continuous improvement of teaching and learning aimed at closing the achievement gap.
- **Appendix A** is a model of a slim curriculum outcomes booklet—a clear statement of what students need to know and be able to do by the end of the year.
- Appendix B is a sampling of short write-ups to teachers after mini-observations and follow-up talks.
   These were written by a principal and assistant principal in New York City who are using the approach described in this book.
- **Appendix C** is a comparison of my teacher evaluation rubrics with Charlotte Danielson's Framework for Learning.

Throughout this book, the word "principal" is used as shorthand for all school-based administrators. Readers should take this to include assistant principals, deans, department heads, and any other professionals who supervise and evaluate teachers.



## The Challenge

### Closing the Achievement Gap

Knowledge powers a global economy that is utterly unforgiving to the unskilled, uneducated young adult.

- Joel Klein, former New York City Schools Chancellor

became principal of Boston's Mather Elementary School late in the summer of 1987, absolutely determined to boost achievement and convinced that supervising and evaluating teachers was at the core of my role as an instructional leader. But had I reflected more carefully on the preceding seventeen years, when I was a teacher, graduate student, and central office administrator, I might have anticipated some of the bumps that lay ahead.

### SUPERVISION AS SEEN BY A ROOKIE TEACHER

Fresh out of college in 1969, I began teaching at Boston's Martin Luther King Jr. Middle School. Supremely ill-equipped to handle a class of twenty-five energetic sixth graders, I had a rough first year. A supervisor from Boston's central office visited several times and was highly critical, so my first exposure to teacher evaluation was one in which my job was on the line. I was one of a number of first-year teachers at the King, and we all regarded this man with fear and loathing. We groused about how the only things he cared about were quiet students, a clean chalkboard ledge, and window shades pulled down at exactly the same height. Disdain for this vision of good teaching was fiercest among those of us who were having the most trouble with classroom discipline. Imagine our glee when students turned the supervisor's Volkswagen Beetle upside down in the parking lot one spring afternoon.

But the supervisor was right to criticize my teaching, and the point was driven home when I invited a professor from Harvard's Graduate School of Education to observe. He sat patiently through a couple of lessons and said afterward that he hadn't seen "one iota of learning" take place. This was not exactly what I wanted to hear, but the comment, from a somewhat more authoritative source, was right on target.

One of the school's assistant principals was assigned to the sixth-grade corridor, and he knew I was struggling. But there were so many other crises in the building that he wasn't able to give me detailed feedback or substantive help.

Somehow I got through the year without being fired—perhaps an acute teacher shortage in Boston helped—and spent the beginning of the summer writing an article vividly describing my experiences (Marshall, 1970, "Law and Order in Grade 6E," published a little later in the *Harvard Bulletin*). After it came out, I received perhaps the most devastating evaluation an idealistic young urban teacher could receive:

Your article clearly shows that whites do NOT belong in Black schools. With all your woes and problems, you forget that the 25 Black students you "taught" have had another year robbed from them (and people wonder why when they become adults they can't "make it" in society). It is unfortunate that you had to "gain your experience" by stealing 25 children's lives for a year. However, Honky—your day will come!

—From one Black who reads the *Harvard Bulletin* 

In my second year, I implemented "learning stations"—a decentralized style of teaching, with students working on materials I wrote myself—and right away things were calmer and more productive. The principal was quite supportive of my unconventional teaching style, even bringing visitors up to my classroom from time to time. But I rarely got any direct evaluative feedback. Did my students learn a lot? I believed they did, judging from weekly tests I created, but I was never accountable to any external standards. These were the 1970s, there was no state curriculum to speak of, and measurable student outcomes weren't part of the conversation. For the school's embattled administrators, the important thing was that there were almost no discipline crises or parent complaints emanating from my classroom.

During these years, I operated very much as a loner, closing my classroom door and doing my own thing. At one point I actually cut the wires of the intercom speaker to silence the incessant schoolwide PA announcements. Here was teacher isolation at its most extreme; if World War III had broken out, my students and I might have missed it.

### **OUT OF THE CLASSROOM**

After eight years of teaching, I stepped out of my classroom to act as the King School's "education coordinator"—a grant-funded curriculum support role that allowed me to work on curriculum improvement but barred me from evaluation because I was still in the same bargaining unit as my colleagues. As I moved around the school, I noticed that the curriculum was highly fragmented, with teachers covering a wide variety of material without a coherent sequence from Grade 6 to 7 to 8, and the quality of teaching varied widely, with no agreed-upon definition of best practice. I saw all this clearly, but my "soft" administrative status prevented me from making much of a difference. After two years as education coordinator, I returned to the classroom, believing that I could have more impact teaching one group of students.

But it wasn't the same. I had definitely been bitten by the administrative bug, and this was reinforced as I pondered a series of *New York Times* articles about an intriguing wave of research on schools that somehow managed to get very high student achievement in tough urban neighborhoods. One prominent exponent was Harvard Graduate School of Education professor Ronald Edmonds, who boiled down the formula for effective urban schools to five variables:

- Strong instructional leadership
- High expectations
- A focus on basics

- Effective use of test data
- A safe and humane climate

A 1979 British study, *Fifteen Thousand Hours* (Rutter, Maughan, Mortimore, Ouston, & Smith, 1979), had a similar message, describing the "ethos" and expectations that made some schools much more effective than others. All the effective-schools research emphasized the importance of the principal going beyond routine administrative functions and being an instructional leader. I began to think seriously about becoming a principal.

The problem was that I didn't have administrative certification, so in 1980, I bid an emotional farewell to the King School, where I had spent eleven formative years, and enrolled in Harvard's Graduate School of Education. I had the good fortune to study with Ronald Edmonds himself, and his searing comment on failing urban schools became my credo:

We can, whenever and wherever we choose, successfully teach all children whose schooling is of interest to us. We already know more than we need in order to do this. Whether we do it must finally depend on how we feel about the fact that we haven't so far [1979, p. 23].

I was raring to go, but during my year in graduate school, the voters of Massachusetts passed a taxlimiting referendum, sending Boston into a budget tailspin and forcing the district to close twenty-seven schools. There was no way I was going to be a principal in the near future, and I prepared to return to the classroom.

Then, through a chance connection, I was recruited to serve on the transition team of Boston's new superintendent, Robert "Bud" Spillane, a forceful advocate of high student achievement and school accountability. He and I hit it off immediately, and I ended up spending the next six years in the central office, first as a speechwriter, policy adviser, and director of curriculum, then, under Spillane's successor, Laval Wilson, as director of an ambitious systemwide strategic planning process. The *Nation at Risk* report of the National Commission on Excellence in Education dominated the national discourse during this period, and I found myself in the thick of Boston's response to the "rising tide of mediocrity" acerbically described in the report.

My central-office colleagues and I did some useful work—we produced a set of K-12 grade-by-grade learning expectations and curriculum tests—but throughout my six years as a district bureaucrat, I felt that our efforts to improve schools were like pushing a string. There weren't enough like-minded principals at the other end pulling our initiatives into classrooms, and we didn't make much of a dent in Boston's abysmal student achievement. I was more convinced than ever that the real action was at the school level, and I longed to be a principal.

### **MY OWN SHIP**

In 1987, I finally got my chance. Laval Wilson put me in charge of the Mather, a six-hundred-student K–5 school with low achievement and a veteran staff. As I took the reins, I believed I was ready to turn the school around after having seen the urban educational challenge from three perspectives: as a cussedly independent teacher, as a student of the research on effective urban schools, and as a big-picture central office official. Now I could really make a difference for kids.

So how did it go? During my fifteen years as Mather principal, the school made significant gains. Our student attendance rose from 89 percent to 95 percent and staff attendance from 92 percent to 98 percent. Reading and math scores went from rock bottom in citywide standings to about two-thirds of

the way up the pack. In 1999, the Mather was recognized in a televised news conference for making the biggest gains in the MCAS (Massachusetts Comprehensive Assessment System, the rigorous statewide tests introduced the year before) among large elementary schools statewide. And in the spring of 2001, an in-depth inspection gave the Mather a solid B+. I was proud of these gains and of dramatic improvements in staff skills and training, student climate, philanthropic support, and the physical plant.

However, these accomplishments came in agonizingly slow increments and were accompanied by many false starts, detours, and regressions. Graphs of our students' test scores did not show the clean, linear progress I had expected. Far too many of our students scored in the lowest level of the 4–3–2–1 MCAS scale, too few were Proficient and Advanced (the top two levels), and our student suspension rate was way too high. Serious work remained to be done. In 2002, I was exhausted and concluded that I had done as much as I could do and it was time to move on. Packing up my office, I hoped that my vigorous young successor would take the school to the next level.

Why weren't Mather students doing better? It certainly seemed that we were pushing a lot of the right buttons, and if the Mather's student achievement had been extraordinary, outside observers would have pointed to a number of "obvious" explanations: my seventy-eight-hour work weeks, the hiring of a number of first-rate teachers, frequent classroom supervision, extra funding and other resources, major improvements to the building and grounds, a daily memo communicating operational matters and research findings to all staff (dubbed the *Mather Memo*), and more. But our student achievement was not extraordinary. Why?

Looking back, I can identify a number of factors that made it difficult for me to get traction as an instructional leader. Teacher supervision and evaluation were the hardest of all, and Chapter Two will describe my struggle to get into classrooms and give teachers meaningful feedback. Others included staff expectations, the school's unique culture, teacher isolation, curriculum fragmentation, poor alignment of teaching and assessment, and unclear goals. Let's examine these challenges (which were hardly unique to the Mather) and an external event that finally began to break the logjam.

### **Low Expectations**

From the moment I arrived at the Mather, I was struck by the staff's unspoken pessimism about producing significant gains in student learning. Teachers had never seen an urban school with really high achievement, were discouraged by the poverty and crime around the school (85 percent of our students qualified for free and reduced-price meals), and had internalized U.S. cultural beliefs about the innate ability level of students like ours. As a result, many staff members saw themselves as hard-working martyrs in a hopeless cause; they loved their students (at least most of them) and did their best, but realistically, high achievement didn't seem to be in the cards. As for the new principal's starry-eyed speeches about the "effective schools" research, teachers were skeptical.

Sensing this ethos, I took a big risk and brought in Jeff Howard, the charismatic African American social psychologist, to explain his "Efficacy" philosophy to the whole staff at an all-day professional meeting in the fall of 1987. Howard held teachers spellbound as he argued that people are not just born smart—they can *get smart* by applying effective effort. He said we could dramatically improve our results by directly confronting the downward spiral of negative beliefs about intelligence and effort. Over lunch, most of the staff buzzed with excitement.

But that afternoon Howard had to leave for another speaking engagement, and the Efficacy consultant he left in charge was peppered with questions from the most skeptical members of the staff. Was he saying that teachers were racist? Was he implying that teachers were making the problem worse? And what did he suggest they do on *Monday?* As the meeting wore on, it was clear that my gamble to unite the staff around a novel approach to higher expectations was going down in flames. As teachers trooped out that afternoon, even those who were sympathetic to the Efficacy message agreed that the day had been a disaster.

In the months that followed, I licked my wounds and took a more incremental approach. In private conversations, team meetings, the staff memo, and clipped-out research articles, I tried to convey the message that higher student achievement was doable at schools like the Mather. I sent small groups of teachers to Efficacy training and eventually brought in one of Howard's colleagues to do a three-day workshop for the whole staff. It was an uphill battle, but work-hard-get-smart beliefs gradually found their way into the school's mission and it became taboo to express negative expectations about students' potential.

### A Resistant Culture

For years, the Mather's staff had been dominated by a small group of very strong personalities, and they did not take kindly to my idealistic approach to urban education or to the fact that I had gone to Harvard (twice!) and had worked in the district's evil central office. The "Gang of Six," as I dubbed them privately (a reference to the Gang of Four, China's maligned leadership team during the Cultural Revolution), began to undermine my agenda with a vehemence that was unnerving. Monthly confrontations with the Faculty Senate, the forum used by the resisters, invariably got my stomach churning. A parody of the *Mather Memo* ridiculing me was slipped into staff mailboxes: "For Sale: Rose-Colored Glasses! Buy Now! Cheap! Get that glowing feeling while all falls apart around you."

I tried to keep up a brave front, but I could not hide my dismay when I heard that on the day of the Efficacy seminar, one of these teachers was overheard to say in the bathroom, "If I had a gun, I'd shoot Jeff Howard dead." At another point, one of these teachers put a voodoo doll likeness of me in the teachers' room and stuck pins in it. Other teachers were so spooked that they didn't dare touch it, and the doll stayed there for several days until a teacher finally had the courage to throw it in the garbage.

Unprepared by my upbringing and limited leadership experience with this kind of behavior, I was sometimes off balance, and every mistake I made became a major crisis ("People are outraged! Morale has never been worse!" said one of the leaders). One such kerfuffle was provoked by the ratings I gave teachers in the initial round of performance evaluations I was required to do in the fall. At this point, Boston's teacher evaluation system had three ratings: Excellent, Satisfactory, and Unsatisfactory. I felt that I hadn't been in teachers' classrooms enough to give them Excellent ratings, and not wanting to devalue the currency, I gave Satisfactory ratings to almost everyone. Although I explained this decision carefully and promised that many ratings would go up when I had time to make more thorough classroom visits, the Satisfactory ratings were taken as an insult by many teachers.

Some of the school's brashest teachers, sensing my weakness and lack of street smarts, went off on me within earshot of others. When I failed to set limits on what could only be described as outrageous and insubordinate behavior, I lost face with the rest of the staff. The "silent majority" secretly wanted me to step up and change the negative culture that had dragged down the school for years, but were so intimidated by the negative few that they remained on the sidelines, which greatly discouraged me. To friends outside the school, I took to quoting Yeats: "The best lack all conviction, while the worst are full of passionate intensity."

Over the next few years, the most negative teachers gradually transferred out—but they had understudies. Every year I battled (not always very skillfully) for the hearts and minds of the silent majority, and

only very gradually did the school develop a more positive culture. How much better things would have been had I possessed the charisma to unify the staff in a quest for higher student achievement! Lacking that, I yearned for "turnaround" powers to deal with teachers who didn't support the mission.

### **Teacher Isolation**

In my first months as principal, I was struck by how cut off Mather teachers were from each other and from a common schoolwide purpose. I understood teachers' urge to close their classroom doors and do their own thing—after all, that's the kind of teacher I had been. But the effective-schools research and my experience in the central office convinced me that if Mather teachers worked in isolation, there might be pockets of excellence, but schoolwide performance would continue to be disappointing.

So I struggled to get the faculty working as a team. I wrote the *Mather Memo* every day and tried to focus staff meetings on curriculum and best practices. I encouraged staff to share their successes, publicly praised good teaching, and successfully advocated for a number of prestigious "Golden Apple" awards for the best Mather teachers. I recruited a corporate partner whose generosity made it possible, among other things, to fund occasional staff luncheons and an annual Christmas party. And I orchestrated a major celebration of the school's 350th anniversary in the fall of 1989 (the Mather is the oldest public elementary school in the nation), fostering real pride within the school and community.

But morale never got out of the subbasement for very long. Staff meetings were often dominated by arguments about discipline problems and, as a young principal who was seen as being too "nice" with students, I was often on the defensive. We spent very little time talking about teaching and learning, and teachers continued to work as private artisans, sometimes masterfully, sometimes with painful mediocrity—and overall student achievement didn't improve.

### Weak Teamwork

Lacking the chops to unite the whole staff around a common purpose, I decided that grade-level teams were a more manageable arena in which to build collegiality. I figured out how to schedule common planning periods for each team (by sending each grade level to specialist classes at the same time), and same-grade teachers began to meet at least once a week and occasionally convene for after-school or weekend retreats (for which teachers and paraprofessionals were paid). A few years later, a scheduling consultant showed us how to create double-period (ninety-minute) team meetings once a week by scheduling art, computer, library, music, and physical education classes back-to-back with lunch. This gave teams enough time during the school day to really sink their teeth into instructional matters.

After much debate, we introduced "looping," with all the fourth-grade teachers moving up to fifth grade with the same students and fifth-grade teachers moving back to fourth to start another two-year loop with new groups of students. Teachers found that spending a second year with the same class strengthened relationships with students and parents—and within the grade-level team—and a few years later the kindergarten and first-grade teams decided to begin looping, followed a few years later by the second- and third-grade teams.

But despite the amount of time that teams spent together, there was a strong tendency for the agendas to be dominated by ain't-it-awful stories about troubled students, dealing with discipline and management issues, and planning field trips. I urged teams to use their meetings to take a hard look at student results and plan ways to improve outcomes, and I tried to bring in training and effective coaches to work with the teams, but I had limited success shifting the agendas of these meetings. In retrospect, I probably would have been more successful if I had attended team meetings and played more of a guiding role, but I was

almost always downstairs managing the cafeteria at this point in the day, and told myself that teachers needed to be empowered to run their own meetings.

### **Curriculum Anarchy**

During my early years as principal, I was struck by the fact that most teachers resisted aligning instruction with a common set of grade-level standards. During my years in Boston's central office, I had worked on nailing down citywide curriculum goals, and I was saddened by the degree to which these official Boston Public Schools expectations were ignored at the other end of the pipeline. While Mather teachers (like many of their counterparts around the country) enjoyed their unofficial academic freedom, it caused lots of problems as students moved from grade to grade. While teachers at one grade emphasized multiculturalism, teachers at the next judged students on their knowledge of state capitals. While one team focused on grammar and spelling, another cared more about style and voice. While one encouraged students to use calculators, the next wanted students to be proficient at long multiplication and division.

These ragged hand-offs from one grade to the next were a constant source of unhappiness. But teachers almost never spoke up to colleagues in the grade just below who had passed along students without important skills and knowledge. Why not? Well, that would have risked getting into some serious pedagogical disagreements that would jeopardize staff "morale" (that is, congeniality). But *not* having those honest discussions doomed the Mather to a deeper morale problem (lack of *collegiality*) stemming from suppressed anger at what many teachers saw as students' uneven preparation for their grade—and lousy test scores that became increasingly important and public as the years passed.

The lack of clear grade-by-grade curriculum expectations was also a serious impediment to my supervision of teachers. When a principal visits a classroom, one of the most important questions is whether the teacher is on target with the curriculum—which is hard to define when no one is sure exactly what the curriculum is! If principals don't have a clear sense of what (for example) second graders are supposed to learn in math and what proficient writing looks like by the end of fifth grade, it's awfully hard to give effective supervisory feedback. And it's impossible for a principal to address this kind of curriculum anarchy one teacher at a time. Supervision can't be efficient and effective until curriculum expectations are clear and widely accepted within the school.

I saw this do-your-own-thing curriculum ethos as a major leadership challenge and tried repeatedly to get teachers to buy into a coherent K-5 sequence with specific objectives for the end of each grade. At one all-day staff retreat in a chilly meeting room at the John F. Kennedy Library overlooking Boston Harbor, I asked teachers at each grade to meet with those at the grade just below and then with those just above and agree on a manageable set of curriculum hand-offs. People listened politely to each other, but back in their classrooms, they made very few changes.

Undaunted, I brought in newly written Massachusetts curriculum frameworks and national curriculum documents, but these didn't match the norm-referenced tests our students were required to take and could therefore be ignored with impunity. When the Boston central office produced a cumbersome new curriculum in 1996, I "translated" it into teacher-friendly packets for each grade level—but once again, these had little impact on what teachers taught. Visiting classrooms, I could comment on the process of teaching but had great difficulty commenting on content.

The lack of coherent learning standards resulted in far too many of our students moving from grade to grade without the skills and knowledge they needed to be successful. As I shook fifth graders' hands at graduation each June, I knew they were better prepared than most Boston elementary students, but we

were sending them off to middle school with major academic deficits. It was not a pretty picture, and I was intensely frustrated that I could not find a way to change it.

### Weak Alignment between Teaching and Assessment

As I struggled to clarify the K-5 curriculum, it occurred to me that perhaps I could use the standardized tests that most Boston students took to get teachers on the same page (*what gets tested gets taught*, I'd been told more than once). The citywide assessment in reading and math at that time was the Metropolitan Achievement Test, given at every grade level except kindergarten, with school-by-school results helpfully published in Boston newspapers. I spent hours doing a careful analysis of the Metropolitan and, without quoting specific test items, presented teachers at each grade level with detailed packets telling what the test covered in reading and math.

Did teachers use my pages and pages of learning goals? They did not. The problem was that the tests teachers gave every Friday (covering a variety of curriculum topics with differing expectations and criteria for excellence) had a life of their own, and I wasn't providing a strong enough incentive for teachers to give them up.

And as hard as it was for me to admit, teachers were not being irrational. The Metropolitan, a norm-referenced test, was designed to spread students out on a bell-shaped curve and was not aligned to a specific set of curriculum goals (Boston's or any other school district's) or sensitive to good teaching (Popham, 2004a). In other words, it was possible for teachers to work hard and teach well and not have their efforts show up in improved Metropolitan scores. Teachers sensed this, and the result was cynicism about standardized testing—and the kind of curriculum anarchy I found at the Mather.

Although my foray into test-based curriculum alignment was unsuccessful, I had stumbled upon an important insight. The key to getting our students well prepared by the time they graduated from fifth grade was finding high-quality K-5 learning expectations and tests that measured them. The problem was that we had neither, and without clear expectations and credible tests, I couldn't coax teachers out of their classroom isolation. For ten years I searched for the right curriculum-referenced tests and tried to clarify and align the curriculum—but until the late 1990s, I wasn't successful. This, in turn, stymied meaningful grade-level collaboration and meant that when I made supervisory visits to classrooms, I was largely flying blind.

### **Mystery Grading Criteria**

Another aspect of the Mather's balkanized curriculum was the lack of agreement among teachers on the criteria for assessing student writing. As is the case in many U.S. schools, the same essay could receive several different grades depending on which teacher read it. The absence of clear, public scoring guides meant that students got very uneven feedback and most teachers lacked the data they needed to improve their classroom methods.

In 1996, the Mather staff made a bold attempt to solve this problem. Inspired by a summer workshop I attended with Grant Wiggins, an assessment expert based in New Jersey, we created grade-by-grade scoring rubrics that described the specific characteristics of student writing at the 4, 3, 2, and 1 level in three domains of writing:

- Mechanics/usage
- Content/organization
- Style/voice

Now our standards for writing were clear and demanding, and we could be pretty sure that the same piece of student writing would get the same scores no matter who graded it. We began to give students quarterly "cold prompt" writing assessments (they wrote on a topic with no help from their teacher or peers) in September, November, March, and June. Teachers scored the papers together and then discussed the results.

This process had great potential. We were scoring student writing objectively; we shared the criteria with students and parents in advance (no surprises, no excuses); we were assessing students' progress several times each year; and teacher teams at each grade were analyzing students' work, giving students feedback, and thinking about best practices for teaching writing.

But for several reasons, this initiative sputtered. Scoring and analyzing tests took too long (often several weeks passed from the time students wrote their compositions to the time we scored and discussed them); our graphic display of the data from each assessment didn't show clearly where students were improving and where they needed help; team meetings fell victim to the "culture of nice" (most teachers weren't frank and honest and didn't push each other to more effective methods); and we didn't involve students in the process of looking at each piece of writing and setting goals for improvement. Without these key elements, our writing initiative didn't bring about major improvements in classroom practice or significantly boost students' performance.

### **Not Focusing on Learning**

As the years went by, I became increasingly convinced that the most important reason student achievement wasn't meeting my ambitious expectations was that we spent so little time actually looking at how much students were learning. The teachers' contract allowed me to supervise classroom teaching and inspect teachers' lesson plans, but woe betide a Boston principal who tried to evaluate teachers based on student outcomes. This resistance was well founded at one level: unsophisticated administrators might be tempted to use norm-referenced standardized tests to unfairly criticize teachers for failing to reach grade-level standards with students who had been poorly taught in previous years.

But not looking at results cuts off teachers and administrators from some of the most useful information for improving teaching and learning. Mather teachers, like their counterparts in other schools, fell into the pattern of teach, test, and move on. The headlong rush through each year's curriculum was rarely interrupted by a thoughtful look every few weeks at how students were doing and what needed to be fixed to improve results.

At one point I asked teachers to give me copies of the unit tests they were giving — not the results, mind you, just the tests. Almost everyone ignored my request, which baffled and upset me. But when I checked in with a few teachers individually, I realized it wasn't an act of defiance as much as puzzlement at why the principal would be making such a request. Most teachers saw their tests as private artifacts that were none of my business.

Perhaps they were also self-conscious about the quality of their tests. (Was he going to look for typos?) Unwilling to push the point and distracted by other issues, I didn't follow up. In retrospect, collecting tests and talking about them with teacher teams might have led to some really productive conversations. If I had taken it a step further and orchestrated conversations about how students performed on the tests, then we really would have been cooking. But I almost never got teachers to relax about the accountability bugaboo and talk about best practices in light of the work students actually produced.