



# 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

I 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 tax-limiting 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.

## THE AHA! MOMENT: STATE STANDARDS AND TESTS

Looking over the challenges I wrestled with in my first decade at the Mather, it's easy to see why we weren't more successful at reaching higher levels of student achievement. I was haunted by the knowledge that with each passing year, the achievement gap between our students and those in more effective schools was widening. But how could we combat the hydra-headed challenges and get higher expectations, create a more positive culture, and convince teachers to work in teams on clearly defined learning outcomes? How could we avoid the Matthew Effect, the Biblical prophecy that hangs over all educators: "To those who have, more will be given, and they will have abundance; but from those who have nothing, even what they have will be taken away" (Matthew 13:12).

Like other struggling schools, we needed outside help—and it finally arrived when Massachusetts introduced rigorous external standards and high-stakes testing (the Massachusetts Comprehensive Assessment System, or MCAS) in 1998. What really got people's attention was that in a few years, students who didn't pass the tenth-grade MCAS tests in reading and math wouldn't get a high school diploma. When that message sank in, things changed quite quickly.

As our fourth graders took the first round of MCAS tests, one highly respected fourth-grade teacher burst into tears at a staff meeting. "No more Lone Ranger!" she exclaimed and pleaded with her colleagues in kindergarten, first grade, second grade, and third grade to prepare students better so that she would never again have to watch her students being crushed by a test for which they were so unready. On that spring afternoon in the school's library, you could hear a pin drop. The teacher's emotional plea shone a bright spotlight on the very problems that had been festering for so many years.

At first, there was resistance to the idea of preparing students for an external test. This wasn't surprising, given the years of working in isolation with idiosyncratic, personal curriculum expectations and contending with standardized tests that didn't measure what was being taught. But when I asked Mather teachers to sit down and take a sampling of MCAS test items, and after we got past some initial push-back ("Mr. Marshall, we're not children. This is a waste of time!"), teachers were impressed. Here's what we concluded:

- Although the tests were hard, they measured the skills and knowledge students needed to be successful in the twenty-first century.
- Success on elementary-grade MCAS tests was an essential stepping-stone to getting a high school diploma.
- It was now possible to align our curriculum to external tests because MCAS items and Massachusetts standards were available online.
- Most of our current students were ill prepared to do well on the MCAS.
- Nonetheless, our kids *could* reach the proficient level if the whole school taught a well-aligned K–5 curriculum effectively over a period of years.

This was just where the staff needed to be in order to take the next steps.

It's worth noting that beginning in 1998, Massachusetts has had high-quality curriculum standards and assessments; a recent study by the Thomas B. Fordham Institute rated Massachusetts materials the best in the nation. Many other states have not been as fortunate, making standards-based reform problematic. The advent of the Common Core State Standards, which are of world-class quality, provides hope for aligning teaching and learning in all states that adopt them. And the advent of a new generation of tests created by PARCC (Partnership for the Assessment of Readiness for College and Careers) and Smarter

Balanced Assessment Consortium may have an impact similar to what we experienced at the Mather School in the late 1990s.

## SLIM CURRICULUM BOOKLETS AND ACHIEVEMENT TARGETS

One problem with the original 1998 Massachusetts frameworks and tests was that they covered only Grades 4, 8, and 10. As Boston curriculum officials mulled over how to fill in the gaps, the Mather staff decided we could do the job more quickly on our own. The state had published “bridge” documents to accompany the Grade 4 MCAS tests, and we set up committees that worked with consultants over the summer to tease back the Grade 4 standards to Grades 3, 2, 1, and kindergarten and up to Grade 5.

That fall, we used the “tease-back” documents to create slim booklets for each grade (about twelve pages long for each level) containing clear learning expectations accompanied by rubrics and exemplars of good student work. Parent leaders helped us scrub the jargon out of our drafts, and our corporate partner printed copies of the booklets for all teachers and parents. The curriculum summaries quickly became drivers for learning in every classroom—and were widely circulated in other Boston schools in what the superintendent at the time referred to as a “curriculum black market.” (See Appendix A for a more recent example of such a booklet.)

Embracing the new Massachusetts standards was enormously helpful in each of the areas we’d struggled with for so long. Grade-by-grade MCAS-aligned targets put an end to curriculum anarchy and focused teacher teams on methods and materials that would maximize student learning, bringing more substance to grade-level team meetings. Although teachers gave up some academic freedom in the process, their isolation from each other was greatly reduced and teams had a common mission.

External standards also helped our staff confront the issue of expectations; having agreed that the new Massachusetts standards were appropriate and attainable (provided there was effective, aligned teaching across the school), we could unite around a relentless push for *proficiency*—a term that acquired special potency when it was attached to the demanding third level on the 4–3–2–1 MCAS achievement scale.

External standards also gave us a more focused mission statement and school improvement plan. Our purpose, we now saw, was to prepare students with the specific knowledge and skills to be proficient at the next grade level, so that fifth-grade graduates would be prepared to achieve at a proficient-or-above level in any middle school. Such a simple and measurable purpose was unimaginable before the arrival of MCAS.

At around the same time we took these steps, Jeff Howard made a successful return visit and helped us agree on a schoolwide achievement target for reading, writing, math, and social competency four years down the road. Grade-level teams then spelled out their own SMART goals (Specific, Measurable, Attainable, Results-oriented, and Time-bound) for that year to act as stepping-stones toward the long-range target (see Chapter Four for more details and samples of both of these). Each year, we updated the SMART goals with higher and higher expectations.

## NECESSARY BUT NOT SUFFICIENT

Ronald Edmonds often said that the existence of even one successful urban school proved that there was no excuse for any school to be ineffective. With this message, Edmonds laid a guilt trip on educators who weren’t getting results, and his stinging rebuke may have jolted some educators into thinking more seriously about improving their schools.

But was Edmonds right that we knew in the late 1970s exactly how to turn around failing schools? Did the correlates of highly successful schools provide enough guidance? Was he fair to thoughtful, hard-working school leaders who were struggling with barriers like those I've described? Was he perhaps a little glib about what it would take to close the gap?

There's no question that Edmonds and his generation of researchers gave us an inspiring vision by showcasing the schools that succeeded against the odds and highlighting the factors that seemed to make them work. It's a tribute to Edmonds and others that the "effective schools" lists they produced have held up so well over the years.

But the early literature did not provide a detailed road map to help a failing school get out of the wilderness, and something else was missing: credible external standards and assessments. Without those ingredients, success depended too much on extraordinary talent, personal charisma, a heroic work ethic, a strong staff already in place—and luck. This allowed cynics to dismiss isolated urban success stories as idiosyncratic and claim that the urban school challenge was fundamentally unsolvable.

That said, Edmonds's extraordinarily important contribution was getting three key messages into the heads of people who care about urban schools:

- Demographics are not destiny: children with disadvantages can achieve at high levels.
- Specific school characteristics are linked to beating the demographic odds.
- We therefore need to stop making excuses, get to work, and learn as we proceed.

Coupled with standards and good assessments, these insights have started us on the way to closing the achievement gap. Recent research on the "90/90/90" schools (90 percent children of poverty, 90 percent children of color, and 90 percent achieving at high levels) by Douglas Reeves, Karin Chenoweth, and others has updated the early research with exemplars of highly effective practice. Visiting these schools is one of the most transformational experiences an urban educator can have.

But turning around failing schools and closing the achievement gap is still extraordinarily difficult. Principals and teachers can have the right beliefs and embrace standards, yet still run schools with mediocre student outcomes. In my years at the Mather and in my work coaching principals and reading extensively since I left the school in 2002, I have become convinced that belief and standards are not enough. To be successful, schools need to radically improve the way they handle four key areas: teacher supervision, curriculum planning, interim assessments, and teacher evaluation—all of which can interact synergistically if they are handled well. The following chapters make the case for a new approach that promises to drive significant improvements in teaching and learning.