19

C H A P T E R

The 90/90/90 Schools: A Case Study

Research conducted at the Center for Performance Assessment on the "90/90/90 Schools" has been particularly instructive in the evaluation of the use of standards and assessment. The research includes four years of test data (1995 through 1998) with students in a variety of school settings, from elementary through high school. Our analysis considered data from more than 130,000 students in 228 buildings. The school locations included inner-city urban schools, suburban schools, and rural schools. The student populations ranged from schools whose populations were overwhelmingly poor and/or minority to schools that were largely Anglo and/or economically advantaged.

One reason that the research in these schools was so productive is that the districts maintained careful records on actual instructional practices and strategies. This allows researchers to investigate associations between instructional strategies and academic achievement results. It is important to acknowledge, however, that these results are only associative in nature. We make no claim that a single instructional intervention can be said to "cause" a particular achievement result. What we can say with a high degree of confidence, however, is that there are some consistent associations between some classroom strategies (for example, performance assessments that require writing) and student achievement in a wide variety of tests and subjects. One final note: We make absolutely no claim that the schools in the study were the beneficiaries of any proprietary "program" or "model" of instruction.

The research literature in every field from pharmaceuticals to education contains too many "studies" that purport to show the effectiveness of treatments that the authors of the research have used. Our role in this investigation is that of journalist and researcher, not of architect of any program or intervention. Hence, we do not claim any credit for improved academic achievement that is rightfully due to the students, teachers, and administrators in the study.

III
PART

Characteristics of 90/90/90 Schools

The 90/90/90 Schools have the following characteristics:

- More than 90 percent of the students are eligible for free and reduced lunch, a commonly used surrogate for low-income families.

- More than 90 percent of the students are from ethnic minorities.

- More than 90 percent of the students met or achieved high academic standards, according to independently conducted tests of academic achievement.

The educational practices in these schools are worthy of notice for several reasons. First, many people assume that there is an inextricable relationship between poverty, ethnicity, and academic achievement. The graph in Figure 19.1 expresses the commonly held belief that poverty and ethnic minority enrollment are inextricably linked to lower levels of student achievement.

Figure 19.1

Common Assumptions About the Relationship Between Poverty, Minority Enrollment, and Student Achievement
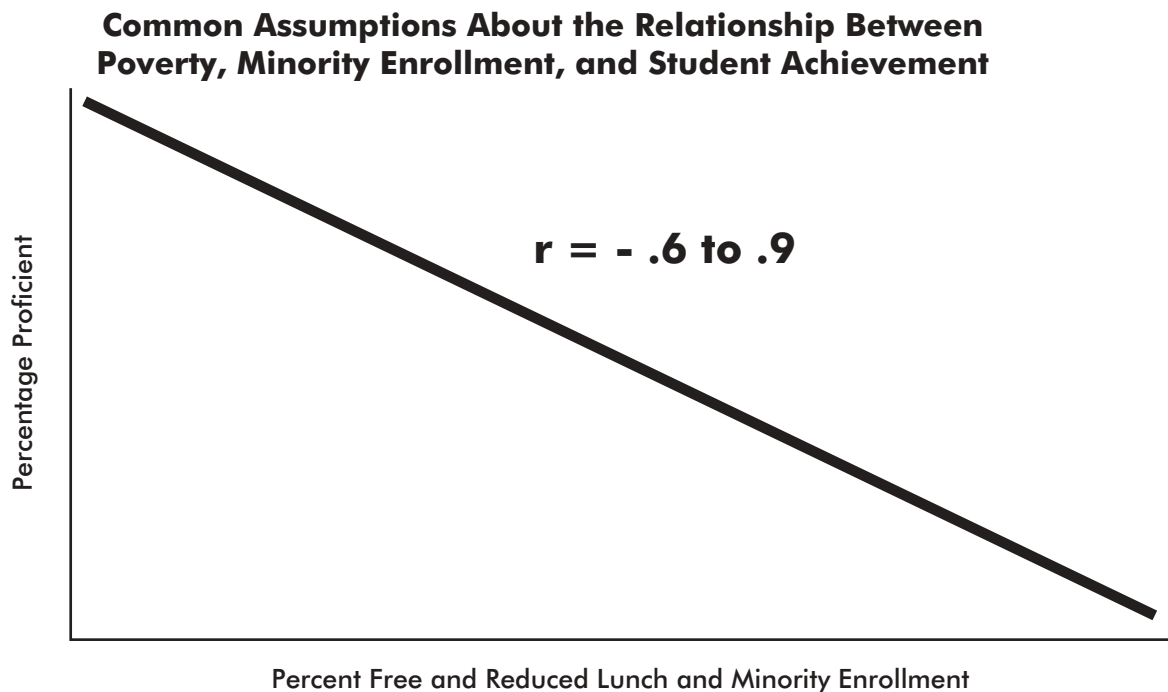
In this chart, the prevailing hypothesis leaves no room for students in the upper right-hand corner of the graph—that is, schools that have high academic achievement coincident with high poverty and high minority enrollments. This is consistent with national observations dating back to the 1960s in which demographic characteristics were regarded as the dominant variables influencing student achievement. In fact, the actual data from the December 1998 Comprehensive Accountability Report of the Milwaukee Public Schools shows a different story. In individual schools, there are striking numbers of students who are

poor and who are members of ethnic minorities who also academically proficient. Throughout the entire system of more than 100,000 students, the relationship between poverty and student achievement is not the postulated -.6 to -.9, but rather a -.2. While the impact of poverty clearly has not been eliminated, the prevailing hypothesis that poverty and ethnic minority status are invariably linked to low student achievement does not conform to the data.

Common Characteristics of High Achievement Schools

Our research on the 90/90/90 Schools included both site visits and analyses of accountability data. The site visits allowed us to conduct a categorical analysis of instructional practices. In the same manner that the authors of *In Search Of Excellence* (Peters and Waterman, 1982) identified the common practices of excellent organizations, we sought to identify the extent to which there was a common set of behaviors exhibited by the leaders and teachers in schools with high achievement, high minority enrollment, and high poverty levels. As a result, we found five characteristics that were common to all 90/90/90 Schools. These characteristics were:

- A focus on academic achievement

- Clear curriculum choices

- Frequent assessment of student progress and multiple opportunities for improvement

- An emphasis on nonfiction writing

- Collaborative scoring of student work

Focus on Academic Achievement

After visiting all of the 90/90/90 Schools, we noticed profound differences between the assessment and instructional practices of these schools and those of low-achieving schools. First and most importantly, the 90/90/90 Schools had a laser-like focus on student achievement. The most casual observer could not walk down a hallway without seeing charts, graphs, and tables that displayed student achievement information, as well as data about the continuous improvement students had made. The data were on display not only in principals' offices, but also throughout the schools. In addition, we saw school trophy cases full of exemplary academic work, including clear, concise essays, wonderful science projects, terrific social studies papers, and outstanding mathematics papers. In short, the 90/90/90 Schools made it clear to the most casual observer that academic performance was highly prized.

The focus on achievement in these schools included a particular emphasis on improvement. The comprehensive accountability system in use by these schools forced every school to identify five areas in which they measured improvement. Although the school could choose the goal from a menu, the common requirement was to focus on a few indicators of improvement in contrast to the typical school improvement plan that contains a large number of unfocused efforts to improve. The focus on improvement is especially important in an environment where many students come to school with academic skills that are substantially below grade level. The consistent message of charts showing weekly improvement from the fall through the spring was, "It's not how you start here that matters, but how you finish." Improvements of more than one grade level in a single year were common, and teachers and administrators paid particular attention to students whose deficiencies in reading and writing would have a profound impact on their success in other subjects. Some students spent as many as three hours per day in literacy interventions designed to get students to desired achievement levels. There did not appear to be any consistency with regard to the intervention programs in use by these schools. Some used Success for All, others used Reading Recovery, while others used the Efficacy Model. Others had no specified program at all, but consistently applied focused intervention for students in need using their own teaching staff.

Curriculum Choices

Such a focus on achievement inevitably leads to curriculum choices, spending more time on the core subjects of reading, writing, and mathematics and less time on other subjects. It is possible, for example, that many of the teachers in these schools did not "cover the curriculum" in the strict sense of checking off objectives from a wide variety of curricular areas. They chose—wisely, we believe—to emphasize the core skills of reading, writing, and mathematics in order to improve student opportunities for success in a wide variety of other academic endeavors later. It is interesting to note parenthetically that, despite their disproportionate emphasis on language arts and mathematics, these schools also significantly out-performed their peer schools on science tests as well. This makes an important point that eludes those who remain committed to a "coverage" model: tests of science, social studies, study skills, and virtually every other subject area are, in fact, tests of reading and writing.

Frequent Assessment of Student Progress with Multiple Opportunities for Improvement

Many of the high-poverty schools included students whose skills were significantly below grade level in academic achievement as they entered the school. The consistent message of the 90/90/90 Schools is that the penalty for poor performance is not a low grade, followed by a forced march to the next unit. Rather, student performance that is less than proficient is followed by multiple opportunities to improve performance. Most of these schools conducted weekly assessments of student progress. It is important to note that these assessments were not district or state tests, but were assessments constructed and administered by classroom

teachers. The consequence of students performing badly was not an admonishment to "Wait until next year" but rather the promise that "You can do better next week."

A frequent challenge to this practice is that students should learn to "get it right the first time." The flaw in such a statement is the implied assumption that the traditional "one-shot" assessment is successful in leading students to "get it right the first time." In fact, when students know that there are no additional opportunities to succeed, they frequently take teacher feedback on their performance and stuff it into desks, back packs, and wastebaskets. Students in this scenario are happy with a "D" and unmotivated by an "F." After all, there is nothing that they can do about deficient performance anyway. In a classroom assessment scenario in which there are multiple opportunities to improve, however, the consequence for poor performance is not a bad grade and discouragement, but more work, improved performance, and respect for teacher feedback. In this respect, the use of teacher evaluation based on assessment scoring guides looked much more like active coaching after which improvement was required, and much less like final evaluation from which there was no reprieve.

Written Responses in Performance Assessments

By far the most common characteristic of the 90/90/90 Schools was their emphasis on requiring written responses in performance assessments. While many schools with similar demographic characteristics employed frequent assessment techniques, many of the less successful schools chose to emphasize oral student responses rather than written responses. The use of written responses appears to help teachers obtain better diagnostic information about students, and certainly helps students demonstrate the thinking process that they employed to find a correct (or even an incorrect) response to an academic challenge. Only with a written response from students can teachers create the strategies necessary to improve performance for both teacher and learner.

In virtually every school we have evaluated, student scores on creative writing are significantly higher than informative and narrative writing scores. As a result, teachers in the successful 90/90/90 Schools placed a very high emphasis on informative writing. They typically used a single scoring rubric to evaluate student writing and applied this scoring guide to every piece of written work. Whether the student was writing a book report, lab report, social studies report, analysis of a sporting event, description of a piece of music, or a comparison of artists, the message was the same: this is the standard for good writing, and there are no compromises on these expectations for quality.

The benefits of such an emphasis on writing appear to be two-fold. First, students process information in a much clearer way when they are required to write an answer. They "write to think" and, thus, gain the opportunity to clarify their own thought processes. Second, teachers have the opportunity to gain rich and complex diagnostic information about why students respond to an academic challenge the way that they do. In contrast to the binary feedback (right/wrong) provided by most

assessments and worksheets, the use of performance assessments that require written responses allows the teacher to diagnose obstacles to student learning. By assessing student writing, teachers can discern whether the challenges faced by a student are the result of vocabulary issues, misunderstood directions, reasoning errors, or a host of other causes that are rarely revealed by typical tests.

The association between writing and performance in other academic disciplines was striking, and this gets to the heart of the curriculum choices that teachers must make. At the elementary level, for example, teachers were faced with a formidable set of curriculum standards in both science and writing. Many of the most successful schools reported that they had to sacrifice time allocated to every other curriculum area except reading, writing, and mathematics. Nevertheless, more than 80 percent of the 135 elementary schools in the study improved in science scores in 1998, compared to 1997. The Pearson correlation between writing improvement and science improvement is striking: .74—a large correlation in virtually any area of social science research. This correlation took place without any changes in the science curriculum and few apparent modifications in teaching methods. I would offer the same caution as provided earlier in the chapter that correlation is not causation. Nevertheless, when two variables appear to behave in such a similar way, it is difficult to escape the conclusion that an emphasis on writing improvement has a significant impact on student test scores in other disciplines, including science.

External Scoring

Another striking characteristic of the 90/90/90 Schools was frequent external scoring of assessments. While many schools continue to rely upon the idiosyncratic judgment of individual teachers for a definition of "proficiency," the high-achieving schools made it clear that no accident of geography or classroom assignment would determine expectations for students. Rather, these schools developed common assessment practices and reinforced those common practices through regular exchanges of student papers. One teacher would exchange papers with another teacher; principals would exchange papers with another school; and in one of the most powerful research findings, principals would take personal responsibility for evaluating student work.

When teachers exchange papers, it is imperative that they have a uniform basis on which to evaluate student work. The degree of agreement among teachers in their use of performance assessment scoring can be measured by "inter-rater reliability." Reliability, when the term is applied to traditional tests, is a measure of consistency. In the case of measuring consistency in scoring, it is simply the percentage of teachers who score an identical piece of student work the same way. If, for example, ten teachers evaluate a piece of student work, and eight believe that the work is "proficient" and two believe that it is only "progressing," then there is an 80 percent reliability rating for that test. This degree of reliability—80 percent—is the target at which teachers should aim as they jointly evaluate student work. It is very unusual (but not unheard of) for that level of agreement to be achieved the first time that teachers jointly score student work. More frequently, there are disagreements

among teachers on the evaluation of student work. These disagreements usually stem from one of two causes. First, teachers frequently use implicit scoring criteria that are not part of the official scoring guide. Examples of implicit criteria include such statements as "He should have written in cursive," or "She knew that she should have included that character in her essay." While these expectations may have been reasonable to these teachers, those criteria did not appear in the scoring guide. It is therefore little wonder that other teachers, who did not share those implicit expectations, failed to mark students down for these failings.

The second cause of teacher disagreement is the lack of clear specifications in the scoring guide itself. Too frequently a disagreement among evaluators leads to an argument rather than to an exploration of how agreement can be achieved through a revision of the scoring guide. "If we change the definition of proficient from this to that, perhaps we could agree on how to mark this paper." Words such as these are the basis of a far more meaningful discussion than, "Of course it's proficient! Don't you see?"

Long-Term Sustainable Results without Proprietary Programs

One of the most powerful findings of the 90/90/90 study is the continuous nature of the success of these schools, even as the poverty of students attending these schools remains intractable. Several of the schools listed below have consistently appeared on the 90/90/90 list, even as students change from year to year, as the effects of poverty grow more onerous, and as parents participating in welfare reform programs are less likely to be at home before and after school. Moreover, these schools are achieving their success without proprietary programs. Let there be no doubt: Our role in this research is as researcher and reporter. None of the 90/90/90 Schools used a specific "program" or any other proprietary model in order to achieve their success. On the contrary, we observed effective teachers and administrators using strikingly similar techniques without the assistance of externally imposed methods of instruction. The techniques used by these schools are replicable, but there is certainly not a need for schools to purchase special textbooks, curriculum materials, or secret information to achieve the level of success enjoyed by these schools.

Non-Proprietary Instructional Practices

In an era in which school leaders appear to engage in a perpetual quest for the magic bullet of educational success, it is noteworthy that none of the 90/90/90 Schools relied exclusively upon a proprietary program to achieve their success. Instead, these schools used consistent practices in instruction and assessment, with support

from local teachers. For those who believe that education remains an interactive process that cannot and should not be "teacher-proofed," these research findings are encouraging. The other edge of this particular razor is that we cannot depend upon proprietary systems to save us. It is the collective work of teachers, students, parents, and leaders that will ultimately lead us out of the present malaise. Every one of the 90/90/90 Schools had academic content standards, but so do many ineffective schools. The distinguishing characteristic of the 90/90/90 Schools was not merely that they had standards, but rather, how the standards were implemented, monitored, and assessed.

Data from the "90/90/90" Studies

A current list of some of the 90/90/90 Schools from Milwaukee, Wisconsin, is provided by the school system in their comprehensive accountability report. Since the publication of the first list in 1998, the number of schools qualifying for the designation has more than tripled. The data were independently verified by Schmoker (2001) in direct interviews with Milwaukee administrators. These schools have graciously hosted hundreds of visitors in the past few years as their successes have become more widely recognized. Researchers and educators should always be willing to share their sources of information and welcome the reviews of colleagues in the field. However, I cannot help but note how profoundly disturbing it is to me that I am frequently requested—demanded is not too strong a term—to produce the names and locations of these schools. In fact, these schools have received significant public attention through the Video Journal of Education, Volumes 802 and 803 (Linton Professional Development Corporation, 1998). Research should, of course, be subject to verification and scrutiny. Nevertheless, I cannot avoid noticing that in my many years of conducting, writing, and reviewing educational research, I have never seen such a demand for "names, dates, and places" accompany the allegation that children who are poor and children of ethnic minority groups perform badly on tests. When *The Bell Curve* (Herrnstein and Murray, 1994) was published with the widely accepted assertion that children who are black and poor perform badly on academic achievement tests, I cannot recall a single instance of demands for the names of students who were subjects of the studies cited. When we have demonstrated that poor and black children perform well, we are inundated with demands for verification. These demands speak volumes about the expectations of children based on their appearance and economic status.

After the original accountability report documenting the 90/90/90 Schools, Milwaukee Public Schools has issued subsequent accountability reports. The findings from these reports are striking. In brief, these findings include the following:

1. Techniques used by the 90/90/90 Schools are persistent. The students are still poor and their economic opportunities have not improved. Nevertheless, more

than 90 percent of the students in these schools continue to meet or exceed state standards.

2. Techniques used by the 90/90/90 Schools are replicable. The first time the district tracked these schools, only seven 90/90/90 Schools were identified. In the most recent report, 13 schools meet the criteria for this distinguished label.

3. Techniques used by the 90/90/90 Schools are consistent. These schools are not lurching from one fad to another. While they differ in some respects with regard to implementation, they are consistent with regard to the following areas of emphasis:

   • Writing—students write frequently in a variety of subjects.

   • Performance Assessment—the predominant method of assessment is performance assessment. This does not mean that these schools never use multiple-choice items. However, it is performance assessment in several different disciplines that local observers have associated with student progress.

   • Collaboration—teachers routinely collaborate, using real student work as the focus of their discussion.

   • Focus—teachers in these schools do not try to "do it all" but are highly focused on learning.

Additional Information on Success in Challenging School Environments

Over the years, I have continued to hear doubts and challenges that poor students can perform well. Indeed, the charge is frequently leveled that comprehensive accountability systems are disadvantageous for poor schools. In fact, systematic research from comprehensive accountability systems allows us to document and celebrate the success of students in these schools. Two additional sources of research on this subject come from strikingly different sources. Casey Carter, author of the "No Excuses" case studies from the Heritage Foundation (1999), provides a conservative viewpoint. The details of these cases are available at www.heritage.org. A politically liberal viewpoint is often associated with Kati Haycock and the Education Trust (1998, 2001). Their landmark research on student success in high poverty schools makes a striking case that these schools are not isolated anecdotes. Indeed, the fundamental finding from the Education Trust studies is that however important demographic variables may appear in their association with student achievement, teaching quality is the most dominant factor in determining student success. It turns out, of course, that teaching quality and subject matter certification are much more likely to occur in economically advantaged schools.

The case made by Haycock and others at the Education Trust is clear: the key variable is not poverty, but teaching quality. While poverty and other demographic variables may be important, they are not determinative in predicting student success. The detailed research from the Education Trust, including an interactive program allowing the user to specify the characteristics of a school and find specific data on comparable high-performing schools throughout the nation, is available at www.edtrust.org.

The consensus of the evidence from very different perspectives is clear: effective teaching and leadership make a difference. The lessons of the 90/90/90 Schools as well as the lessons of other studies provide convincing evidence that accountability systems, properly designed, can provide a wealth of information for those desiring to find the keys to improved achievement for all students.

Using the 90/90/90 Practices to Improve Achievement and Close the Equity Gap

Researchers and practitioners must always confront the gap between theory and reality, between anecdote and evidence. "Sure it worked there," the skeptics say, "but our kids are different." The ultimate test of the 90/90/90 research is whether it is sustainable and replicable. Simpson (2003) provides compelling evidence that the practices of the 90/90/90 Schools can be applied in a diverse urban environment with similar results:

> Like the city, Norfolk Public Schools, the first public school system in Virginia, has seen its fortunes go up and down. It's an urban district that serves a diverse population: 67 percent of students are black and 28 percent are white. More than 65 percent of students qualify for free and reduced-price lunches.
>
> - 100 percent of our schools met the state benchmarks in writing in all grades tested.
>
> - 100 percent of our high schools met the state benchmarks in chemistry.
>
> - 100 percent of our middle schools are fully accredited in earth science.
>
> - 100 percent of our middle and high schools showed positive trends in reading, literature, and research.
>
> Also, our schools reduced the achievement gap between white and black students in third, fifth, and eighth grades, with both groups continuing to improve. They decreased disciplinary actions by 15 percent, the number of long-term suspensions by 14 percent, and the number of expulsions by 66 percent. In addition, we have two "90/90/90 schools." These are schools with more than 90 percent of students eligible for free and reduced-price lunch, more than 90 percent are minority students, and more than 90 percent of

students met high academic standards on the state's Standards of Learning tests. (Simpson, 2003, pp. 43-44).

At the beginning of the 2002-2003 school year, I examined the accountability reports of each of the schools in Norfolk, Virginia, and conducted numerous site visits and interviews. In particular, I wondered if the buildings that experienced gains of 20 percent or more in their academic achievement in language arts, mathematics, science, and social studies were significantly different than their counterparts in other schools. The schools with the greatest gains were not similar demographically, as they included high-poverty and low-poverty student populations. The financial support, staffing patterns, union agreements, and central office support were similar for all schools. Therefore, neither the demographic variables of students nor the external variables of funding and labor agreements could explain the extraordinary differences between the schools. The keys to improved academic achievement are professional practices of teachers and leaders, not the economic, ethnic, or linguistic characteristics of the students. The Norfolk accountability system revealed striking similarities to other research on the characteristics of successful schools. Although surely there are many other traits shared by effective organizations of all types, the Norfolk Accountability System provided an insight into measurable indicators that were linked to the largest gains in student achievement. These characteristics also make clear that successful accountability is not the exclusive domain of the "Department of Accountability" in the central office, but rather is a responsibility shared throughout the system on many levels. The observations made on the basis of this inquiry are strikingly similar to observations I have made in other school systems over the course of several years. The following paragraphs highlight the nine characteristics that distinguished the schools with the greatest academic gains.

The Impact of Collaboration

First, the schools devoted time for teacher collaboration. This was not merely an exercise in idle discussion nor at attempt to get along in a friendly and collegial fashion. Rather, collaboration meetings were focused on an examination of student work and a collective determination of what the word "proficiency" really means. At first, teachers identified wide variations in their opinions and were alarmed to see how differently they evaluated the same piece of student work. In the course of many sessions, the most effective schools made time for collaboration very frequently and in some cases did this every day. Where does the time come from for effective collaboration? None of these schools had extra money in the budget or more hours in the day. Rather, they used the time that they already had with an intentional focus on collaborative scoring of student work. For example, the principals made their faculty meetings "announcement-free zones." Rather than drone through a laundry list of announcements (with inevitable comments and controversies), their rule was that the transmission of information would always be in writing. This allowed time formerly devoted to faculty meeting announcements

to be dedicated to collaboration. The principals were literally on the same side of the table as their faculty members, with faculty members who were experienced in collaborative scoring taking turns facilitating faculty meetings. The other source of time for collaboration was professional development meetings. Rather than presentations by outside staff developers, a significant degree of the professional development time was allocated to collaborative scoring. These educators knew that collaboration is hard work. Moreover, they understood that it is a skill acquired over time. Hence these remarkably effective schools did not have a "collaboration day" or a "collaboration workshop" but rather made the collaborative scoring of student work a part of their regular routine.

The Value of Feedback

Second, the schools with significant improvements provided significantly more frequent feedback to students than is typically the case with a report card. Emulating their most successful colleagues in music and physical education, teachers provided feedback in real time. They knew that a basketball coach does not provide hints on an effective jump shot nine weeks after an error, nor does a great music teacher note the improper position of the violinist's left hand weeks after noticing the mistake, but rather coaches and musicians provide precise and immediate feedback. In some cases, teachers took a triage approach, providing traditional report cards to successful and self-directed students, while providing weekly reports on their progress to students who were struggling. Their approach to feedback was consistent with Robert Marzano and his colleagues whose meta-analysis of research on student achievement revealed that feedback had a profound impact on student achievement, provided that the feedback was timely, accurate, and specific (Marzano, Pollock, and Pickering, 2001). The emphasis that these teachers placed on accuracy in feedback was remarkable. Unlike the "positive distortion" that clouds so much classroom feedback (Foersterling and Morgenstern, 2002), teachers with large gains were committed to feedback that was consistently accurate, with student performance compared to unambiguous expectations.

The Impact of Time

Third, the schools with large gains made dramatic changes in their schedule. Although they had the same budget, state requirements, teacher's union contract, and other restrictions as other schools in the system, the schools with large gains made remarkable schedule changes. At the elementary level, they routinely devoted three hours each day to literacy, with two hours of reading and one hour of writing. At the secondary level, they routinely provided double periods of English and mathematics. This was not a shell game in which they used the block schedule to double up some times but cut back on English and math in other times, but rather represented a genuine increase in instructional hours of math and English. The

essential nature of instructional time is hardly a new idea, yet in an astonishing number of schools, the schedule is revered more than the Pledge of Allegiance, Constitution, and Magna Carta combined. To break the mold in student achievement, these schools discovered, they had to break the schedule. It is interesting that this commitment to time for literacy instruction occurred in a state in which social studies and science content examinations were required. These teachers and principals did not change the schedule to over-emphasize literacy because they disregarded science and social studies, but rather because they knew that literacy was essential for success in every content area.

Action Research and Mid-Course Corrections

Fourth, teachers engaged in successful action research and mid-course corrections. In many of the schools with the greatest gains, their school accountability plans were not static documents set in concrete before the beginning of the school year, but dynamic and flexible guides. They asked the central office for permission to change goals and strategies that were not effective and start new ones that held promise, even during the school year. Moreover, these faculties and leaders learned from one another. An illustration of their commitment to the application of action research is the use of word walls at the secondary level. Because both the school improvement data and the instructional techniques associated with those improvements are transparent in a system of holistic accountability, the teachers who had achieved great things with students were subject to being questioned by colleagues throughout the system about their success. When in earlier years, elementary educators reported that significant improvements in vocabulary and reading comprehension results were associated with the implementation of word walls, the secondary science and social studies educators decided to adopt the idea. They created walls with words containing essential science and social studies vocabulary, sometimes associated with vivid visual images, and used those vocabulary words throughout the year. In other examples of effective action research, teachers replicated one another's writing rubrics, interdisciplinary assessments, and student motivation practices.

Aligning Teacher Assignments with Teacher Preparation

Fifth, principals made decisive moves in teacher assignments. Some writers have argued that when test scores are down, the entire school should be reconstituted and the entire faculty dismissed. In my observations, however, principals have made impressive gains by reassigning teachers to different grades within the same school. Consider what has happened to the curriculum—particularly in the fourth, fifth, and sixth grades—over the past decade. There has been an enormous growth in the

complexity of the curriculum, particularly in math and science, with an accompanying set of assumptions about the undergraduate curriculum of the teachers responsible for those grades. Those assumptions have sometimes been wildly inappropriate. When the fourth grade curriculum requires an understanding of algebra and scientific inquiry and the teacher's undergraduate preparation does not include those subjects, there is a challenge that will not be solved with a one-day staff development course in academic standards. The teachers whose undergraduate backgrounds fail to match the standards are not bad people nor are they unprofessional educators. Rather, their preparation is better suited to a different grade level. Effective leaders know that they should seek not to "fix" the person, but rather find a job (and accompanying set of standards) that best meets the teacher's abilities and backgrounds. By making decisive moves in teacher assignments, these principals saved the careers of some teachers and dramatically improved the achievement of their students.

Constructive Data Analysis

Sixth, successful schools included an intensive focus on student data from multiple sources, and specifically focused on cohort data. They were less interested in comparing last year's fourth grade class to this year's fourth grade class (which are, in most instances, different children) and more interested in comparing the same student to the same student. Their most important questions were not, "Is this year's class different from last year's class?" but rather:

- "What percentage of a group of students is proficient now compared to a year ago?"

- "What percentage of our students have gained one or more grade levels in reading when we compare their scores today to their scores a year ago?"

- "Of those students who were not proficient a year ago, what percentage are now proficient?"

- "Of those students who were proficient a year ago, what percentage are now advanced?"

In brief, these teachers compared the students to themselves rather than to other groups of students. This analysis allowed them to focus their teacher strategies on the needs of their students and not on generic improvement methods.

Common Assessments

Seventh, the schools with the greatest improvements in student achievement consistently used common assessment. This is a dangerous recommendation to

consider in an era in which the most frequently heard complaint across the educational landscape is that students are over-tested. To be sure, many students are over-tested; but they are under-assessed. The distinction between testing and assessment must be clear. Testing implies an end-of-year, summative, evaluative, process in which students submit to a test and the results—typically many months later—are used by newspapers and policy makers to render a judgment about education. By the time the results are published, they are ancient history in the eyes of the student and teacher. Contrast this to the best practice in assessment, in which students are required to complete a task and then very soon—within minutes, hours, or days—they receive feedback that is designed to improve their performance. Effective assessment is what great music educators and coaches routinely provide to their students. Moreover, great educators use assessment data to make real-time decisions and restructure their teaching accordingly. The track coach, for example, does not use the previous year's data to make decisions about assembling relay teams or selecting students to compete for the state finals. Rather, the most recent data available is far more important than the final results from the previous year. Similarly, the data from last quarter on a school-based assessment is far more helpful than the data from last year's test. Common assessments also provide a degree of consistency in teacher expectations that is essential if fairness is our fundamental value. Although individual teachers must have discretion on a day to day and hour to hour basis to teach, re-teach, and otherwise meet the needs of individual students, they do not have the discretion to presume that their students "just can't do it." The use of a common assessment for each major discipline allows for a combination of daily discretion and independence by teachers, while preserving a school-wide commitment to equity and consistency of expectations.

The Value of Every Adult in the System

Eighth, these remarkably successful schools employed the resources of every adult in the system. In holistic accountability systems, we can explore the extent to which professional development is distributed among all adults in the system. In a few remarkable cases, for example, there is profound respect for every employee, including bus drivers and cafeteria workers. The respect for these employees is evidenced by their inclusion in professional development opportunities in classroom management and student behavior. Leaders recognized that the student's day does not really begin in the classroom, but on the bus or perhaps during free breakfast. By committing their systems to consistency in the education and behavior of adults, these leaders ensure that every adult leader, from the bus driver to the food service employee to the classroom teacher is regarded as a significant adult leader in the eyes of students. The language concerning student behavior, sanctions, and rewards, is consistent and the results are impressive. Concomitant with gains in student achievement, these schools witnessed dramatic improvements in student behavior, including a reduction of bus misbehavior and disciplinary incidents outside the classroom.

Holistic accountability (Reeves, 2001) reviews allow a consideration of other extraordinary performances, including those by school nurses, library/media center specialists, school secretaries, custodians, counselors, psychologists, security guards, and many other unsung heroes whose exceptional efforts are disregarded in the typical accountability report. While holistic accountability does not provide a cookie-cutter approach to school success, it does reveal the remarkable impact of every adult in the system on student achievement.

Cross-Disciplinary Integration

Ninth, there is explicit involvement of the subjects that are frequently and systematically disregarded in traditional accountability systems—music, art, physical education, world languages, technology, career education, consumer and family education, and many other variations on the these themes. Analysis of holistic accountability data reveals that the involvement of these seemingly peripheral subjects in academic achievement is neither serendipitous nor insignificant. Rather, there is a deliberate strategy of involvement in these subjects in the improvement of academic results for all students. A few examples will serve to illustrate the point. Teachers meet to review student achievement data at a deep level, including the sub-scale scores. The discussion is not that "math scores are low" but rather that "the sub-scales reveal that we need to work in particular in fractions, ratio, and measurement." This leads the music teachers to develop activities in which musical rhythms reveal the relationship of whole-notes, half-notes, and quarter notes. Art teachers work on perspective and other representational art that makes explicit use of scale. Physical education teachers allow students to choose to run either a millimeter or a kilometer, and when they make the wrong choice, it is a lesson most students remember well.

In a striking example of collaboration in Norfolk, the teachers in music, art, and physical education collaborated to teach a social studies unit about African studies and the nation of Mali, the home of many of the students' ancestors. Using dance, literature, vocabulary, geography, history, song, and other engaging activities that crossed disciplinary boundaries, the teachers took the Mali unit out of the shadows of the final week of school and infused it throughout the school year. It is hardly an accident that these students also displayed astonishing improvements in their performance on state social studies tests.

Other Urban Success Stories

Norfolk is hardly an isolated example of success in urban school systems. In Indianapolis, Indiana, the Wayne Township Metropolitan School Corporation is among many that has demonstrated that academic improvement is compatible with high percentages of minority and poor students in the student body. In St. Louis,

Missouri, Dr. Chris Wright and her colleagues have led successful initiatives in both Riverview Gardens and Hazelwood school districts. Now, under the leadership of Dr. John Oldani and Dr. Dennis Dorsey of the Cooperating School Districts of St. Louis County, these techniques are having an impact throughout the St. Louis area. In Los Angeles County and Orange County, California, urban, suburban, and rural school systems are collaborating to create significant gains in student achievement.

The Wayne Township results are particularly interesting, as they represent not only an example of successful accountability, but also the ability of a complex urban school system to replicate the success of other systems. The Wayne Township experience demonstrates that holistic accountability is not merely the result of idiosyncratic case studies, but rather the result of systematic replication of best practices from within and outside a school system. The demographic characteristics of Wayne Township might be those of any urban system, with 26 different languages spoken by the students, free and reduced lunch enrollment as high as 80 percent in some schools, and minority enrollment increasing in a number of schools to the point that a majority of students are from minority ethnic backgrounds in some buildings. What is unusual, however, is the relentless focus of this school system on collaboration, academic standards, and nonfiction writing at every level. In particular, the years from 1999 through 2003 represent an extraordinary effort to augment the state's accountability system with a district-based holistic accountability system. In addition to the state tests, the district administers pre- and post- tests for every student in the fall and spring of each academic year. For the year ending in June of 2002, every single school made significant gains in mathematics and language arts. In addition, the schools with the highest poverty levels made the greatest gains, perhaps because those schools displayed the most intensive focus on changing schedules, instructional practices, building-level assessment, and leadership. It was therefore no surprise that when the state tests were administered in the fall of 2002, every building displayed significant growth, but those buildings with the highest poverty levels displayed the greatest growth in academic achievement. These gains exceeded 20 percent in the case of several schools within the district.

Without a constructive accountability system, these results might be passed off as the temporary reaction to test preparation resulting from pressure from state authorities. The facts contradict such a presumption. Every school in Wayne Township tracked specific practices in leadership and teaching. In the case of those schools with the greatest gains, there were common assessments on a monthly or quarterly basis. In addition, faculty meetings and staff development sessions were routinely devoted to collaborative scoring of student work. Each of the schools had common scoring rubrics so that there were consistent descriptions of what the word "proficient" means in practice. Following the lead of the district, each school embraced the use of "power standards" so that teachers were able to focus on a few of the most important standards rather than every single standard established by the state. This is among the most important observations of this holistic accountability study: higher test scores resulted not from mindless test prep and frantic coverage of

every standard, but rather from the thoughtful application of the most important standards to creative and engaging teaching strategies.

It was noteworthy that the schools that had the greatest gains did not eliminate special area courses, such as music, art, physical education, and technology. Rather, these courses were explicitly a part of the academic preparation of every student. In schools with the highest gains, each teacher in the special areas was given the standards in mathematics and language arts in which students needed the greatest amount of help. Each of these teachers incorporated some of those language arts and math standards into their daily lessons.

Finally, the principal was personally involved in the evaluation of student work. The building leader regularly met with students and parents to discuss student achievement in specific terms. Moreover, the principals personally administered common assessments every month in language arts and math. By giving up faculty meetings, the principal helped to provide additional time for collaborative scoring of student work. The principal also encouraged every teacher to display proficient and exemplary student work in a highly visible manner. The result of these displays was that every student, parent, and teacher had a clear and consistent understanding of what the school-wide scoring rubrics meant in practice.

The Impact of Holistic Accountability on Equity

As impressive as the improvements in academic achievement were in Wayne Township, the gains in equity were nothing short of extraordinary. Figure 19.1 showed the typical negative relationship between poverty and student achievement. The more likely a school is to have high percentages of poor and minority students, the less likely the school is to have a high proportion of the students achieve academic proficiency.

The line extending from the upper left to the lower right shows that as the percentage of students in poverty (as defined by those eligible for free or reduced lunch) increases, the achievement (as measured by test scores) decreases. This relationship is not perfectly negative (-1.0) but it is substantial in most national research, ranging from -.6 to -.9. The prevailing assertion in more than four decades of research on the topic is that variables such as student poverty account for 90 percent or more of the variation in student test scores (Marzano, 2003). If we stop with a consideration of Figure 19.1, then these prevailing assertions will carry the day. The accountability evidence, however, suggests that there are specific teaching, leadership, and curriculum strategies that will mitigate the impact of poverty.

Figures 19.2 through 19.5 indicate that the negative relationship between student poverty and student achievement is not a certainty. Although the grade 6 language arts scores are disappointingly negative (- .35), in both grades 3 and 6, the relationships between poverty and achievement are far lower than is the case

nationally, and in three out of four examples, the relationships are almost flat. In other words, this school system has demonstrated that the relationship between poverty and student achievement can be negligible.

Relationship Between Poverty and 3rd Grade Language Arts Achievement

r = .03

Percent Proficient

Percent Free and Reduced Lunch

Relationship Between Poverty and 3rd Grade Mathematics Achievement
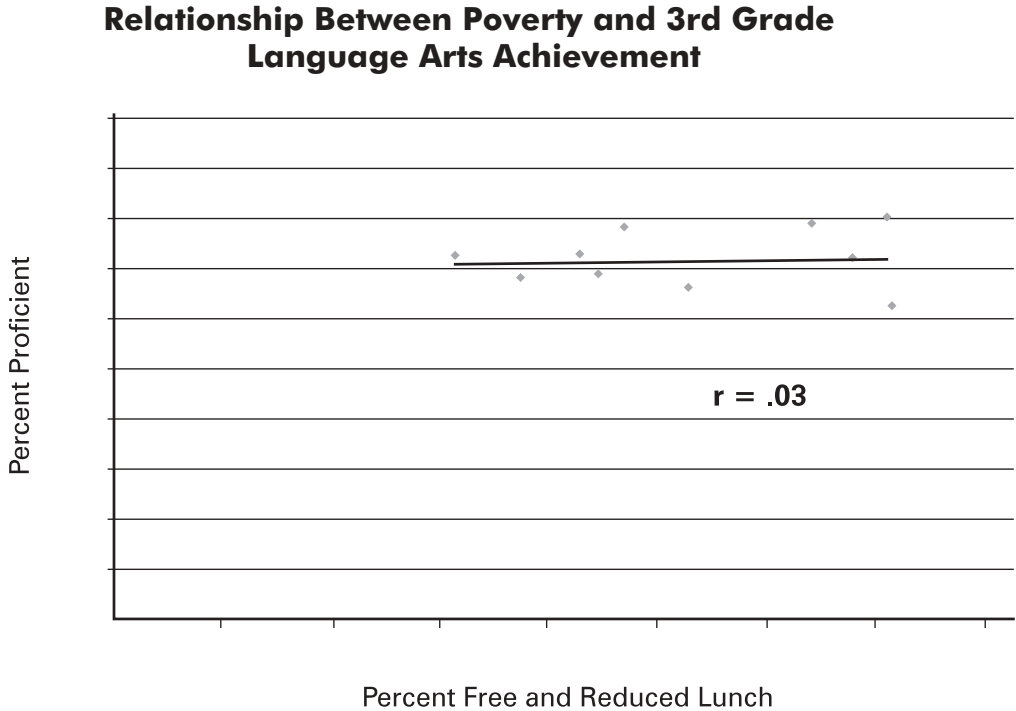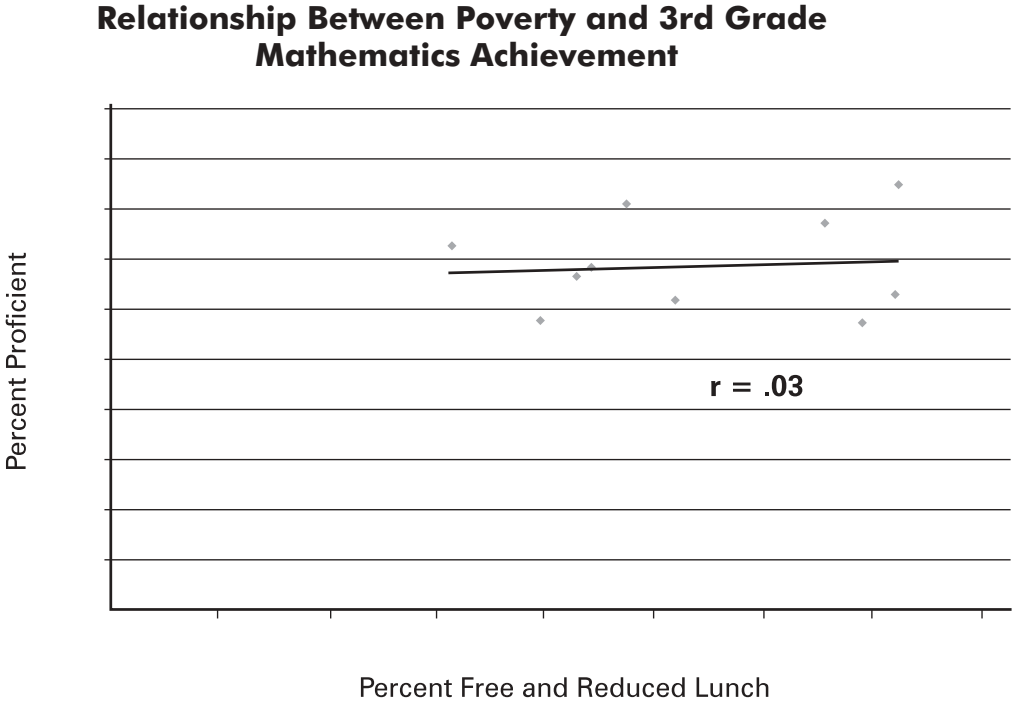

r = .03

Percent Proficient

Percent Free and Reduced Lunch

Equity Need Not Be A Dream

The Wayne Township experience demonstrates that equity need not be a dream. Every single building in the district—elementary through high school—achieved one of the following two equity indicators: The difference between students eligible for free and reduced lunch and the average was less than 10 percent, or the difference between the largest minority group of students and the average was less than 10 percent. These data points are totally consistent with the improvements in equity in Milwaukee, Freeport, Riverview Gardens (St. Louis metropolitan area), and others.

While no one disputes that poverty, linguistic differences, and culture can be important variables influencing student achievement, the research is clear that variables in teaching, curriculum, and leadership are profoundly important. In fact, these variables, that teachers and leaders can control, are more influential over student achievement than the intractable variables of poverty, culture, and language.
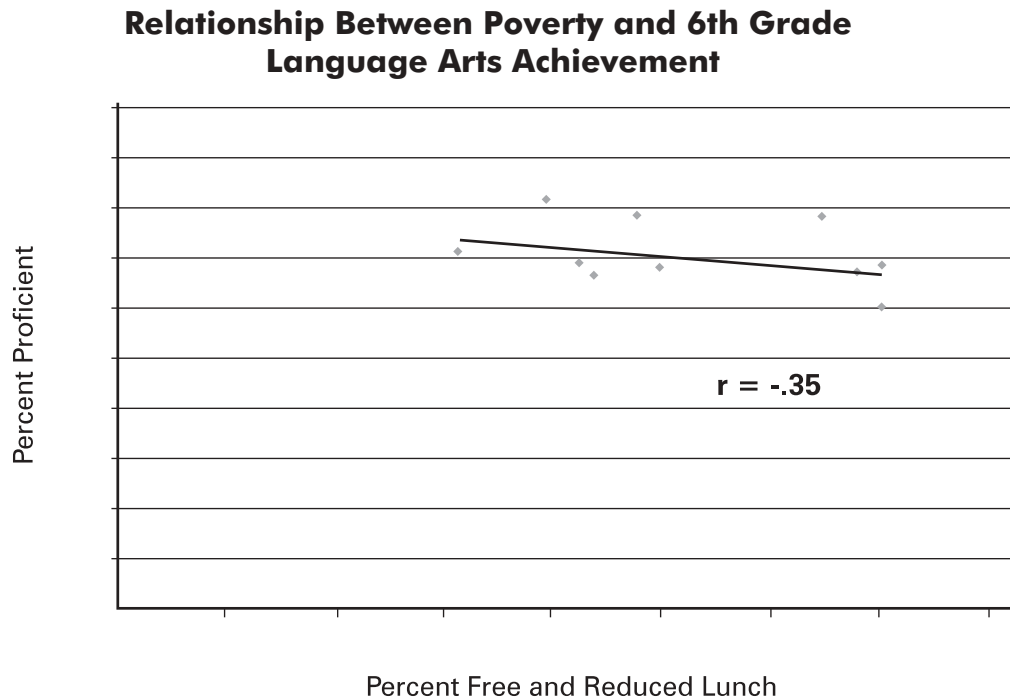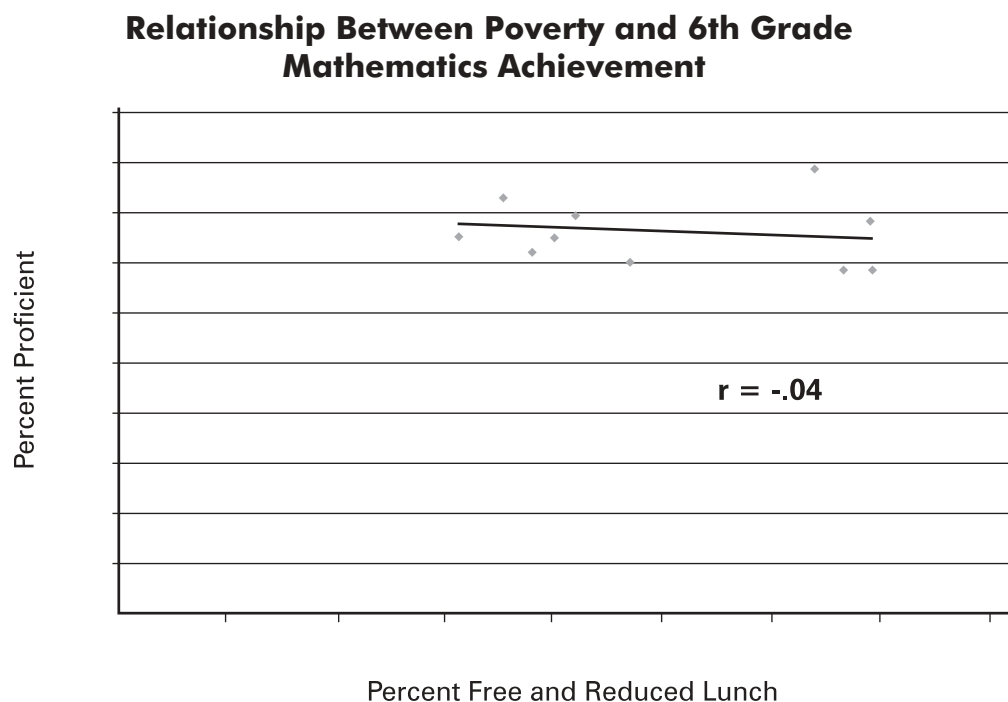
Figure 19.4

Relationship Between Poverty and 6th Grade Language Arts Achievement

r = -.35

Percent Proficient

Percent Free and Reduced Lunch

Figure 19.5

**Relationship Between Poverty and 6th Grade
Mathematics Achievement**

r = -.04

Percent Proficient

Percent Free and Reduced Lunch

Critics, Cynics, and Urban Education Success

We must take a few minutes to address the inevitable critics who appear to be constitutionally unable to believe that a success story in urban education exists. Whenever I share results such as those in Norfolk, Wayne Township, Milwaukee, Riverview Gardens, Freeport, or other successful urban schools, critics inevitably roll their eyes and allege that this surely must be a flash in the pan, the product of a frenzy of test preparation rather than sustainable reform. Others have claimed that the results must be due to the exclusion of under-performing children on test day. Still other critics claim that the students and teachers must be engaged in a massive cheating conspiracy. Others take issue with the methodology of the research, particularly if careful research controls (such as mobility and attendance) are used. The presence of those controls inflates achievement, the critics charge. After all, the studies reflect students who actually attend school. Of course, the absence of those controls would lead to charges of sloppy research. Either way, the critics find a way to ignore the continuing pile of research, of which my studies represent only a few pebbles. Marzano (2003) has assembled the most impressive evidence, using meta-analytic techniques that indicate the importance of teaching, curriculum, and leadership relative to poverty and ethnic identity. Demographic characteristics are relevant, but the preponderance of the evidence indicates that these characteristics are not destiny when it comes to academic achievement. The following is a brief consideration of challenges that I have heard made to the 90/90/90 research:

The only measure of success in this study is test scores, and there are better ways to assess student achievement. Test scores are a way, but by no means the only way, to assess student achievement. It is interesting that one of hallmarks of the 90/90/90 Schools was an unwillingness to tolerate annual state or district tests as the sole measurements of achievement. These schools consistently elevate the importance of classroom-based, teacher-made tests that are collaboratively scored and used to provide immediate feedback to both students and teachers. From a research and policy perspective, however, it is necessary to have some consistent data in order to understand student achievement. While accountability should indeed be a holistic endeavor with multiple assessments of achievement, common tests of literacy and mathematics are useful to evaluate student achievement over time. Finally, the best accountability systems, including the one used in the original 90/90/90 research, included a balance of state, district, and school-based measures. Moreover, it included a narrative report from each school, providing a balance of qualitative observation and quantitative data.

The excessive time devoted to reading means less time for science and social studies. This is true. Schools in the study were required by state law to take science and social studies tests, yet they made a deliberate trade-off to devote more time to reading comprehension and nonfiction writing, even if it meant that they had fewer hours of social studies instruction. This trade-off was wise for two reasons. First, their scores in social studies and science did not decline, but increased. One can speculate that it might have had something to do with the improved ability of students to read and understand the questions on the social studies and science tests. Second, our interviews of social studies and science teachers at the secondary level revealed their nearly unanimous conviction that the key to greater success in those disciplines at the secondary level was not more social studies and science instruction in elementary school, but students who could enter secondary school able to read on grade level. A substantial body of research (Foersterling and Morgenstern, 2002; Klentschy, Garrison, and Amaral, 2000) supports the teachers in this conviction.

The controls for attendance and mobility provide a positive bias for 90/90/90 Schools. This is not true. The accountability system provided "two-column" reporting for students in order to display the impact of mobility and attendance. In one column, the report shows the results for all students, and in the next column it shows the results for those students who were continuously enrolled during the school year. For attendance, the "all student" number was separated from the results for those students who attended school at least 90 percent of the time. These controls were made for all schools, not just the 90/90/90 Schools. Therefore, a parallel comparison was made to high poverty, high minority schools for students with good attendance and continuous enrollment, but who did not have the success of students in the 90/90/90 Schools. This is just good research design. In pharmaceutical research, we compare patients who receive the medicine (the experimental group) to those who receive a placebo (the control group). The research is only useful if those in the experimental group really take their medicine. If we are studying the impact of certain strategies in curriculum, teaching, and educational leadership, our research is of questionable value if we analyze the effects on students who were not present

for the curriculum, teaching, and leadership strategies. Finally, it was noteworthy that the schools that had high mobility (as defined by more than 80 percent of students taking the spring test not enrolled in September) and also high achievement, had strikingly similar characteristics to the 90/90/90 Schools, with an emphasis on writing and collaboration.

The 90/90/90 Schools used expensive programs, such as Success for All. This is not true. Some schools used Success for All, and others did not. This makes emphatically clear that the brand name alone of a literacy program is not the predictor of success, but rather the professional practices employed by teachers and leaders in the building. In fact, some Success for All schools had high results, while others had poor results. It was the replicable professional practices, not particular programs, that were associated with student success.

The effects are transient and dependent upon a particularly effective principal and faculty. This is not true. The effects are sustainable, with some schools maintaining this designation through different principals and high faculty turnover. The effects are replicable, with schools in other places (where there is also high turnover and teacher inexperience, particularly in high poverty schools). In the words of one teacher in the original study, "nobody volunteered to come to this school." Nevertheless, their collaboration, focus, and professional practices delivered results.

Conclusion

Perhaps the most compelling argument against any research about success in high poverty schools is the observation that there are cases where teachers are doing all of the right things, and yet student achievement remains low. There are no magic potions to deliver improved student achievement. The best that researchers and policymakers can do is to examine the preponderance of the evidence and draw appropriate conclusions. When a jury is presented with the evidence in a court case, it rarely has a perfect data set with unquestionable research. Rather, the jury confronts conflicting information, including information with errors, uncertainties, and differing interpretations. From this mix, we ask twelve people of good will and common sense to draw an appropriate conclusion based on the preponderance of the evidence.

The 90/90/90 research and the other evidence offered in this article fall far short of perfection. It does, however, contribute to the larger body of evidence that, in its totality, suggests useful strategies for high poverty schools. Moreover, in any research project, we must recognize that perfection is not an option. Rather, we can only choose among the errors that we commit, and attempt to minimize the risk of our errors. From a research perspective, we must choose between the risk of confirming a hypothesis that is not true and the risk of failing to confirm a hypothesis that is true. In the case of the professional practices recommended in this

article, we also have two potential errors. One error is the replication of these practices, including an increase in our commitment to literacy, nonfiction writing, and collaboration, and the subsequent discovery that the students really did not need all of that extra work after all. What is the risk of this strategy? Excessively literate students? Teachers who collaborate too much? The other error is the failure to act while we search for perfection or persist in a state of disbelief. Risks attendant with such delay will be debilitating for another generation of students. I do not claim that the 90/90/90 research and its many counterparts in the literature are perfect. I only suggest that the risks of this research being wrong are minimal. The risks if the research is correct and ignored are grave.

Questions for Discussion

1. What implications do the research findings on "90/90/90 Schools" have for your school or district? How would you implement these in your school or district?