Chapter 16

TEST PREPARATION FOR SUCCESSFUL SCHOOLS

Before anything else, preparation is the key to success.

—Alexander Graham Bell

Issues and Themes

The level of success in life achieved by children, the careers of educators, and even the value and stability of entire communities is now linked to high-stakes tests required in our schools. Under the provisions of NCLB, each school within a district (except parochial and private) must publish a report card reporting among other things the disaggregated scores on mandated statewide assessment tests. There can be community-wide feelings of angst and disappointment when local schools fail to meet required adequate yearly progress goals. Avoiding the need to inure the inevitable reforms and structural changes that would be brought by the community is a powerful source of motivation for administrators. The stress of this pressure is felt by every teacher in every classroom. Under these circumstances, it is inevitable that cheating the system will occasionally occur. Such cheating violates the laws of most states as well as the ethical canons of the professional associations.

The stress is even more palatable for the students who feel they are at jeopardy for disappointing their parents and losing face within the community

of other children. This pressure on children even extends to state policy. Michigan now awards high school graduates a $2,500 college scholarship if they pass the state assessment tests. This is not a small issue for many working-class families.

The influence of the high-stakes tests in the lives of children is now part of our popular culture and is elaborated in children's books and dramatically presented by the feature filmmakers of Hollywood. As our society as a whole is in favor of school and student accountability, there is no doubt that this pressure on our children will be with us for the foreseeable future.

Both parents and schools have used early (kindergarten and first grade) retention to give children a developmental advantage when they enter third grade. A number of other marginally ethical practices have been employed by educators to improve the scores of children on high-stakes tests. Some of these involve peer pressure, and others amount to little more than bribes for improved work.

A number of strategies exist for improving the assessment scores of students that are both effective and ethical. These involve the administrative leadership style, teacher inservice education, and the school's staffing and scheduling practices. Within the classrooms there are also specific instructional strategies that are effective in improving scores on the mandated assessments.

The length of the school year and the length of the school day have been examined for their influence on achievement outcomes. While little has come from these efforts, related studies of class size and school plant architecture have been shown to correlate to assessment test scores. Also, there are a number of ways that the organization and administration of the assessment tests can relate to the outcome.

These approaches to testing and test preparation may soon be the concerns of a different era as we move toward a totally wired school where all testing and remediation is facilitated online. Each year the number of states using an online method for student assessment is increasing, and by the final year of the original goals of the NCLB Act (2014), these methods will be universal.

Learning Objectives

By reading and studying this chapter you should acquire the ability to do the following:

- Explain the sources of stress on educators that are part of the high-stakes testing movement.

- Provide examples from children's literature and the popular culture of the stressful impact of high-stakes testing on children.
- List test preparation activities that are not legal and/or ethical.
- Discuss the question of staffing and teacher assignment to the outcome of high-stakes tests.
- Explain how the leadership style of a school administrator can play a role in the assessment test scores of students.
- Design a community outreach program for a school that could work to improve assessment test scores.
- Describe pre-reading and reading strategies that have been found to relate to higher scores on mandated assessment tests of reading.
- Outline strategies that can be employed to improve the writing scores of students.
- Provide strategies that can be used to improve the mathematics test scores for elementary and middle-school students.
- Explain the distinction between "teaching to the test" and "teaching for the test."
- Describe the research on the relationship between the length of the school day and school year on assessment test scores.
- Discuss the link between leadership style, school climate, and achievement outcomes.
- Describe the research consensus on the relationship between class size and student achievement.

ACCOUNTABILITY ANGST

The mind that is anxious about the future is miserable.

—Santana

One goal of the NCLB Act is to provide an open and transparent accountability system for all public schools. This involves the publication of "report cards" for every school building, school district, and state education department. At the school building level this process includes reports of the number of children at each level of proficiency on each of the required tests. It also requires that these scores be disaggregated for each minority and other identified group of children. Other items such as the number of "highly qualified teachers" who are employed to work in the school, and the student attendance and dropout rates, are also included.

Cost of Failure

From 2002 to 2005 it was necessary for adequate yearly progress to be shown by each group of students in each grade level toward meeting the ultimate goal of every student being proficient in every tested subject by 2014. In 2005, that was changed by U.S. Department of Education Secretary Margaret Spellings so that progress could be made on the basis of some grade levels, not all grade levels. That policy shift may have been in answer to the many middle-class suburban schools that found they missed meeting the AYP goals and were labeled as "needing improvement" (Tracey, Sunderman, & Orfield, 2005). The label is almost always presented to the public by local media as "failing." The general public has no way to discern anything from such headlines beyond the fact that their expensive and heretofore highly regarded suburban school system is a failure. In 2006, there were a total of 1,750 schools that had a 5-year history of failing to meet the AYP standard and were in the process of being reorganized or closed (Feller, 2006). That same year 26% of all schools, a total of 22,873, failed to meet the AYP target.

The tweaking of the rules by the Department of Education notwithstanding, respected school administrators have had their reputations sullied by media reports of the systemic failure of local schools. Entire communities can feel the impact when the scores on the mandated testing program are reported in the local press. Kurt Landgraf (2005), president of ETS, has cautioned against reading too much into assessment test scores. His caution has been ignored by the media. It had to be expected that real estate agents would follow the reports of test scores, knowing that families with young children shop for homes where the school system has a good reputation (Lloyd, 2005; Van Moorlehem, 1998). Today, even divorce lawyers have found the scores of interest when they negotiate for the custody of children.

These reports, and the inevitable public response, have put significant strain on the whole enterprise of public education. Pressure on school leaders is imparted to the teaching staff, who can find themselves surrounded by nervous children worried about the high-stakes tests, and building principals anxious for their school to score well on the next state-issued report card. As an increasing number of states move to using "value added" assessments with their annual tests, individual teachers who are less effective than their peers will become easier to identify. The future portends only more assessments and an ever-increasing emphasis on measurable outcomes.

At the personal level, children can also feel the stress that high-stakes assessments have on their lives. In 2007, approximately 100,000 nine-year-old children failed third grade and were not promoted to the fourth grade with their peers because of low test scores. In the 27 states where high school

graduation is contingent on passing an exit assessment, 30% to 40% of the senior class who are non-White minorities and about half of that number of the White students will fail and not earn a standard high school diploma.

Even those high school students who have good grades and who score well on the state tests still face college admissions. College admission looms as a major source of concern until the "fat" letter full of admission material arrives in the April mail.

Case in Point (16a)

The perception of an admission crisis is expanded out of proportion in the hallways and lunchrooms of high schools when students begin to receive the dreaded "thin letter" of the wait list. School counselors often unknowingly add to the stress and feeling of urgency by posting letters of acceptance on the bulletin boards of high schools. There once was a time, not so long ago, when the goal of going to college was to get an education. What few guidance counselors or parents know is that there are only about 50 colleges in the United States that send out more letters of rejection than they do letters of acceptance (Steinberg, 2002). Some parents have been described as deeply engaged in the study of the details of college catalogues, attempting to ferret out the subtle differences between the top-ranked private liberal arts colleges in the Northeast (Cohen, 2006). Often these parents experience more concern about college admissions than do their children. Yet, the pressure is real, and teenagers are expected to stand out and somehow appear to be stellar candidates for the college admissions office. To this end, families have been known to move to communities with less competitive high schools, making it possible for the student to have a better GPA and higher class rank. Adolescents have been known to volunteer to work with community groups and agencies, not out of a sense of duty or altruism but to make their college applications look better (Matthews, 2005).

Popular Culture

Popular culture has also picked up on the stress that results from mandated assessments and high-stakes tests. Children's literature has begun to play on this theme with fictional books such as Judy Finchler's (2000) *Testing Miss Malarkey*; Pansie Hart Flood's (2004) *It's Test Day, Tiger Turcotte*; Edward Bloor's (2004) *Story Time*; Andrew Clements's (2004) *The Report Card*; and Nancy Lieberman's (2004) *Admissions.* Even the fifth book of the

wildly popular children's series by J. K. Rowlings (2003), *Harry Potter and the Order of the Phoenix,* finds the hero, Harry Potter, under stress to achieve a passing score on the "Ordinary Wizarding Levels Examination (OWLs)."

Hollywood has not missed this trend either. These movies are designed for an adolescent market and have common themes (viz., that youngsters must face the evil forces of the testing and assessment; then, only through their cleverness, can adolescents rise above these impediments and become successful). This is the theme of *How I Got Into College,* released in 1989 by Twentieth Century Fox (Shamberg, Cantillon, & Holland, 1989). In *Stand and Deliver,* the 1988 release from Warner, a class of inner-city students is accused of cheating on an AP examination and are subjected to a grueling retest by factotums of the ETS Company (Law & Mendez, 1988). In 2004, Paramount Pictures released *The Perfect Score,* which tells the story of a cabal of high school seniors who conspire to steal the answers to the SAT (Birnbaum & Robbins, 2004). Disney released *Recess: School's Out*, a story of a radical educational reformer who wants to cancel recess to have time to have students work on better test scores (Suzuki, Swuz, & Sheetz 2001). In 2005, Hart Sharp Video released the film *Admissions,* a fictional account of a high school graduate flubbing in her college admission's efforts (Vait & Painter, 2005). In 2002, Columbia Pictures released the film *Stealing Harvard*, the story of an uncle who turns to crime to pay the tuition to Harvard for his niece (Grazer & McCulloch, 2003). Also in 2002, Paramount released the film *Orange County*, a comedy about the admission of a surfer into Stanford University (Rudin & Gale, 2002). A romantic comedy about admissions into Princeton, *Cinderella Story*, tells the story of the life of a young woman applying for admissions into an Ivy League college (Rachmil & Rosman, 2004).

During the summer of 2005, the NBC network ran a 6-week series using a "reality TV" format that focused on college admissions. This series, *The Scholar,* had 10 high school seniors compete for financial aid to attend the colleges of their choice.

Cutting Corners and Cheating

In light of the amount of pressure on students and educators, it is not surprising that corners have been cut and incidences of cheating have occasionally been reported (Burney, 2006). Even entire states have been described as "gaming the system" (Pascopella, 2007a). While unusual, these incidences are widespread, involving an estimated 1% of the educators in our schools (Donsky, 2005). It is difficult to know how common this type of fraud is, because teachers usually do not talk to outsiders about what their colleagues are doing.

The type of cheating that occurs most often involves extra time for slower students and letting faster students work ahead on the test, answering questions in sections before they should. Other inappropriate strategies have involved having aids posted on the walls of the classroom (e.g., multiplication tables, vocabulary words). More scandalous problems have involved changing the answers of students once the test has been completed, teaching the specifics of what is on the test, and helping students while the test is in progress.

These problems seem to erupt every spring during testing season and have the potential to erode public confidence in the integrity of educational reforms under the NCLB Act. The reality is that in most states no one is even looking for unusual patterns of scores and cheating. One company, Caveon Test Security, has established a niche for itself by using statistical models to analyze school scores and even student answer sheets to detect possible cheating. Starting in 2005, the schools of Texas, Ohio, and North Carolina had their scores scrutinized by consultants. The states of Nevada, Louisiana, and Mississippi have hired in-house staff to perform such reviews (Patrick & Eichel, 2006). Meanwhile, Pennsylvania, South Carolina, and Illinois include in the test developer's contract the requirement of test reviews.

Caveon found that there were 699 Texas schools that could be identified as having unusual and hard to explain shifts in student scores from one year to the next. However, the most common way that these schools were identified was by an analysis of answer sheets, which found that too many students made the same errors and got the same items correct (Benton, 2006). Many of these anomalies occurred with the schools of Houston, where a number of administrators were subsequently demoted and teachers reprimanded or fired (Hacker & Parks, 2005). The full audit of all of Texas's schools in 2006 indicated a high probability that 1 school in 11 reported fraudulent scores on the statewide assessment.

James Popham (2006) suggested that cheating by teachers occurs because the state-mandated assessments are both correlated to the socio-economic status of the test takers and are instructionally insensitive. Thus, no matter how much effort a teacher puts into instruction and how hard he or she tries, the telling factor in student outcomes is predicated on how wealthy the test takers' families are.

Information about one company specializing in test security can be viewed at www.caveon.com/press6-3-05.htm.

Because most states provide a window of time when the schools can administer state assessments, teachers from one school have been known to e-mail or call friends in other buildings and provide a "heads-up" as to what the test contains. Parents have also been known to call relatives with children

attending other schools to pass on test information, such as writing prompts and any other items their children may remember about the test.

LEGAL AND ETHICAL REQUIREMENTS

For the past 30 years, committees of the professional associations and learned societies of education have developed and published guidelines for ethical use of tests and assessments (Impara, 1996). The most recent edition of the *Code of Fair Testing Practices in Education* provides the following two relevant statements for educators administering high-stakes assessments: "Follow established procedures for administering tests in a standardized manner" and "Protect the security of test materials, including respecting copyrights and eliminating opportunities for test takers to obtain scores by fraudulent means" (Joint Committee on Testing Practice, 2005, p. 26). What this statement of ethical practice does not provide is any method of enforcement.

The introduction of high-stakes testing programs by the states changed all this by introducing an era of statute-enforced testing procedures. Now unethical testing practices are also illegal. Statutes of the various states now provide specific investigative steps and penalties for a breech of ethics. These penalties can involve forfeiture of a teaching and/or administrative license or even more serious criminal penalties. To accompany the new high-stakes assessments, the states have written prescriptive testing ethics codes and standards for practice.

MARGINAL ETHICS

While many of the methods noted above are clearly unethical and deserving serious sanctions, many other methods to influence test scores of children are questionable, but legal.

Parental Strategies

Parents have been known to transfer their child to a parochial or private school at the end of second grade. Those schools are not required to use a high-stakes test with students. After a year or two the child is quietly transferred back into public education. This strategy is not successful in states where there are tests for promotion at several grade levels. Parents can also opt to home-school a child for a year or two to avoid the jeopardy of the high-stakes test in third grade. Another parent strategy that is occurring with

BOX 16.1 "Testing Code of Ethics," as Adopted in West Virginia

WEST VIRGINIA DEPARTMENT OF EDUCATION

126CSR14
TITLE 126
LEGISLATIVE RULE

SERIES 14
WEST VIRGINIA MEASURES OF ACADEMIC PROGRESS (2340)
§126-14-8. Investigation of Security Violation.

8.1. Any written complaint and/or oral report alleging a violation of West Virginia Measures of Academic Progress testing security or copyright infringement shall be reported immediately to the county superintendent of the county (or organization) in which the violation(s) occurred. The county superintendent shall immediately contact the West Virginia Department of Education who will advise the county about the procedures to follow regarding any investigation.

8.2. All complaints will be investigated whether reported by a named individual or anonymously to insure test security for all students and reliability of school results.

8.3. An investigation must occur in a timely and efficient manner and shall be:

8.3.1. jointly conducted by the designated staff of west Virginia Department of Education and the county superintendent (or the designated staff) of the county in which the alleged violation occurred, or

8.3.2. at the request of the county superintendent, conducted by the staff of West Virginia Department of Education, solely.

8.4. The investigation must, at a minimum, consist of personal interviews with the reporter, if known; the individual(s) against whom the allegation(s) is/are filed; and any other individual who may have knowledge of the alleged incident.

8.4.1. The investigation may also consist of any other methods and review of the circumstances deemed pertinent by the investigators within the bounds of the law.

8.4.2. The investigating team must take immediate steps to protect the rights of the complainant, students, teachers, administrators and other personnel including the individual(s) against whom the allegation(s) is/are filed pending the completion of an investigation of testing security or copyright violations.

(Continued)

BOX 16.1 (Continued)

8.5. Upon completion of the investigation, written findings and recommendations based on all the facts and surrounding circumstances must be submitted immediately to the Office of the State Superintendent of Schools with a copy to the county superintendent and shall include:

8.5.1. a determination of whether any action or incident constitutes a violation of testing security procedures or copyright infringement as follows:

  a. testing security procedures or copyright infringement was not breached, or

  b. testing security procedures or copyright infringement breach could not be determined, or

  c. testing security procedures or copyright infringement was breached and include recommendations for

    A. invalidating the test scores of students/classes/school/county,

    B. retesting of students with the equivalent form of the test with the county assuming the cost of both purchasing and scoring the equivalent form, if testing window has not closed,

    C. re-aggregating the test data for valid test results for students and school,

    D. using the equivalent form in the next year's administration.

8.5.2. Any taped conversation from the investigation shall be transcribed and a copy shall be forwarded to the county superintendent upon written request following the submission of the recommendations above.

8.6. Upon receipt of the written findings and recommendations of the investigation team, the county superintendent shall take any appropriate employment action and so advise the State Superintendent of Schools, in writing, setting forth any recommendations as follows:

8.6.1. no action was taken by the county board,

8.6.2. punitive action was taken by the county board,

  a. suspension by county board,

  b. termination by county board, or

  c. other.

8.6.3. action against teaching license taken by State Superintendent of Schools.

8.7. Upon review of the investigation team's written findings/recommendations, the State Superintendent of schools shall take appropriate action against those found to have violated test security procedures or copyright violations.

SOURCE: WV Dept. of Education 126CSR14, Title 126 Legislative Rule, Series 14, WV Measures of Academic Progress (2340) §126-14.8. Investigation of Security Violation.

greater frequency involves holding the child an extra year in a private kinder-garten before enrolling him or her in first grade in a public school. It is likely that the school systems will see the grade level cohorts being a bit older than in the past, as this and similar methods are used more frequently.

Private tutoring and coaching for the tests have had an upsurge in pop-ularity. Private schools that coach for the exam have become commonplace. Some of these private test cram centers have even taken the name of their Japanese counterparts, *jukus*.[1]

School Strategies

Some school administrators have taken a similar approach to improving the average test scores of students. By selecting slower children for grade retention while they are in first or second grade, it is possible to give them an extra year to catch up before they face the high-stakes test in third grade. This strategy can be initiated by parents who wish to give their child the advantage of an extra developmental year over their peers. The professional associa-tions oppose grade retention for this purpose, making it a sketchy approach for a school to adopt (National Association of School Psychologists, 2003).

Another marginal school-based strategy for raising school scores on the state test was uncovered by an economist, David Figlio (2003). He found that during the times during the year when the high-stakes tests are given, and during the days set aside for make-up testing, there is a differential pattern of punishment administered by school disciplinarians. Figlio found that low-performing students who are caught violating a school rule during the testing period are significantly more likely to be punished with an extended out-of-school suspension than are others. Also, he found that students who perform well on tests who are caught in violation of a school rule during the time when the mandated tests are occurring are given less harsh punishments and usu-ally are included among those taking the assessments. This process only occurs during the years when the high-stakes tests are administered.

While it is not ethical to teach children how to answer the items on the high-stakes test they will face, an alternative strategy involving teaching old (released) test items is becoming commonplace.[2] Several states (California, Florida, Maryland, Massachusetts, Ohio, Texas, and Virginia) release some or all of their high-stakes tests a year after they have been used. An ethical ques-tion arises when these old tests are included as part of the curriculum. The direct instruction of last year's high-stakes test is clearly "teaching to the test" and not teaching the knowledge base and cognitive skills that make up the state's learning standards. This process circumvents the whole purpose of the high-stakes test, which is to assess the extent to which children have

learned those skills and areas of knowledge specified in the standards. Instead, the mandated high-stakes assessment becomes a measure of how well children can remember items and test-gaming strategies.

Rewards have always been used by parents and educators to encourage children to do their best work. Recently, schools have employed a number of different rewards to motivate children to perform better on mandated assessment tests. These have included tickets to concerts, admission to professional sporting events, and free meals at local restaurants (Belluck, 2006; Knight, 2005; Lou, 2007; Pakkala, 2006; Woods, 2007). Occasionally school administrators have even used foundation funding to present savings bonds to students who show improvement on the assessments. The principal of an upper-middle-class high school that had a severe parking shortage devised a novel motivational reward. The highly coveted and very limited student parking permits were awarded on the basis of scores on the state assessment test. One school in Florida ran a cram course on nine Saturdays to get at-risk children ready for the state's mandated test. They awarded perfect attendance at the Saturday juku with a new iPod (Crouse, 2006). Other Florida high schools have provided limousine rides and prom tickets to juniors who do well on the mandated high school examinations (Bailey, 2007). Another Florida high school used various fundraisers to raise $5,000, which was used to purchase 100 MP3 players for all the juniors that did well on the state mandated examination, while an elementary school that was given a grant of $2,000 to buy school supplies used the money to buy 400 five-dollar gift cards at a local store. These were used as rewards for the children for good test scores (Ehrlich, 2007). A high school raised $18,500 to buy a new Scion car that was given as a raffle prize. The raffle was only open to those scoring at the proficient level (or above) on the state's examination.[3] The principal of an elementary school and several devoutly religious members of his faculty returned to school the night before the high-stakes tests started and blessed with holy oils the desks where children would sit to take the test the next morning (Wood, 2007).

Case in Point (16b)

Perhaps the most unusual of these reward programs for children occurred in Florida. In some of the elementary schools of the City of Gainesville, the third grade students who performed best on the state assessment, the FCAT, got to stand in a chamber filled with swirling money and grab as many handfuls of $1 and $5 bills as they could in 15 seconds. This unusual project was cosponsored by a local bank.

Merit Pay

It was not until the 1960s that the traditional unequal pay scales of school systems were replaced by a single pay scale for all teachers employed in a district (Johnson, 2000). Early pay systems favored high school teachers at the expense of elementary school teachers.[4] That single pay scale is now being questioned. Critics feel that highly effective teachers should receive extra compensation above and beyond the common pay scale, in the form of merit pay or a bonus.

In some regards, the merit pay offered by a number of school systems to their teachers is similar to reward programs noted above for students. That is to say, merit pay is believed to be a way to reinforce quality teaching and thereby improve the education of all children. Additionally, it is believed that merit pay will help recruit and retain high-quality teachers. The argument that proponents of merit pay make is that the present system of having one standard pay scale for all teachers only encourages mediocrity and discourages the best and most motivated educators. Opponents of merit pay point out that merit awards are summative statements that are made only after a teacher demonstrates meritorious teaching. The concern is that merit pay provides neither a formative evaluation nor developmental steps designed to improve those teachers who are not at that meritorious level. A second concern is that teacher incentives do not address the central problems of education. It can be argued that teachers have not been withholding their best efforts, waiting for the day when they get bonus pay for better instructional efforts (Hershberg & Lea-Kruger, 2007). If classroom teachers are to be paid for performance, they should be given far more control over the instructional environment then they have today.

Merit Pay vs. Signing Bonus

A second argument that merit proponents make is that teachers of some subject areas are very difficult to find. Teachers of science, mathematics, modern language, and special education will be more likely to work in a system that offers a salary incentive based only on their disciplines. In big league sports this is referred to as a signing bonus. Critics of merit pay point out that such salary incentives amount to so little money as to be trivial to experienced teachers and would not compare favorably with the pay packages offered in the private sector to recent science and mathematics graduates.[5] Yet, there is a danger of any number of perverse effects when the incentives are substantial (Firestone, Monfils, Schorr, Hicks, & Martinez, 2004).

A number of names have been used with the concept of merit pay for teachers, including pay for performance and incentive pay. There are four ways merit pay can be awarded to a teacher. These can fit into a 2-by-2 matrix:

AWARDED TO WHOLE SCHOOL	AWARDED TO INDIVIDUAL
One-time bonuses	One-time bonus
Permanent salary increases	Permanent salary increase

Merit Pay as a Unit Reward

When the decision is made to award a whole school, there is still the question as to whether all the teaching staff will have an equal share or whether the principal will be empowered to decide who on the staff is to be given an award and how much that will be. Most school administrators do not wish to be the person who must determine merit levels and provide justifications for the decision on each teacher's merit. Elementary school administrators fear that the merit pay decision will destroy the within-grade level cohesion and morale. High school principals with a professional staff of 200 certified teachers and specialists see the job as a bureaucratic nightmare of paperwork. Most teachers do not want administrators making the merit award decision, fearing awards would go to "obsequious boot lickers" (Lieberman, 2000). They are concerned that all creativity within the classroom and all constructive criticism of the policies within a school would be stifled. These concerns notwithstanding, there is evidence that teachers have begun to have more positive opinions of the practice of awarding merit pay based on test scores (Dillon, 2007b).

Merit Pay Problems

Another concern that most teachers have is that merit pay will be predicated on the scores of mandated assessments. This type of an evaluation may also include a value-added assessment based on one of the various regression models now available. In either case, the fear is that a low-achieving class one year could affect a teacher's livelihood. Another concern has been expressed by those teachers who are not directly engaged in classroom instruction, such as librarians, media specialists, art and music teachers, and reading specialists. Other teachers in grades where there is no test (e.g., kindergarten and first grade) are also concerned with the evaluation system. Their efforts will not appear on the assessment test scores. At the secondary level, 10 or more teachers are involved in the education of any one student in a year's time. The

problem for high school administrators is to unravel how effective each of these educators is and make an appropriate award to them for their efforts.

Merit Pay in Play

The school systems of several large cities (e.g., Cincinnati) have working models for merit pay in place. When state departments of education require merit pay for all school personnel, local school boards often must foot the bill.[6] In Florida, a statewide merit pay plan has been in place and generally ignored since 1998, and Colorado began a performance pay plan in 2006. In 2006–2007, the State of Texas initiated a $250 million program of merit pay incentives (Hacker & Stutz, 2006). The first year is a pilot year, and 2007–2008 merit pay plans are being initiated in over 1,000 schools serving mostly low-income children. The awards are in the form of one-time bonuses that range between $3,000 and $10,000. With the mean teacher salary in Texas (2006–2007) at $42,000 a bonus of $10,000 represents a before-tax boost of 23.8%. It is generally hoped that this plan will lure better teachers into the schools of impoverished communities. The first round of awards went to 7,900 teachers in the city of Houston. Almost all of these merit awards were given to teachers of science and mathematics and averaged $1,850.[7] The awarding of these merit bonuses produced a loud chorus of complaints from the community who felt that the selection criteria were wrong (Tonn, 2007).

To date there is no definitive empirical evidence that most large-scale merit pay plans have been effective in bringing about educational improvement (Firestone, Monfils, Hayes, et al., 2004; Jacobson, 2006). However, research based on 1992 data has indicated that there may be a small positive effect associated with merit pay in school applications (Viadero, 2007, January). This contrary finding may be a statistical artifact, but it does indicate that incentive programs seem to work best in schools populated by children of poverty. It needs to be emphasized that when surveyed, effective teachers report that their pay is not the most critical issue in the decision to remain employed by the school system. Teachers want a work environment that is replete with educational materials, where they are respected by their administrators and have good relationships with colleagues and students (Laitsch, 2007; Organization for Economic Co-operation and Development, 2005).

SUCCESSFUL SCHOOLS

There are a number of ethical and legal activities that educators can do to improve student scores on the mandated assessment tests. Some of these

16-Wright-45489.qxd 11/20/2007 3:48 PM Page 510

510   PART V   NO CHILD, TEACHER, OR SCHOOL LEFT BEHIND

involve leadership and school building climate, others are instructional in
nature. A third approach to improving student outcomes involves data
management.

A number of the steps that can be taken to improve the proportion of
students who are "proficient" on the mandated assessments have nothing to
do with students or their instruction. The state departments of education can
change the rules and make more schools and children seem to be successful.
One example of this is lowering the bar, or cut scores, that defines what is
necessary to be proficient. Another is to change the rules related to student
enrollment. In 2006–2007, the State of Illinois redefined which students'
scores from the state's test would be counted in determining a school's ade-
quate yearly progress (AYP) level. To have a student's score counted in the
determination of a school's AYP level, he or she had to be enrolled from
October 1 until the Illinois test was given in the spring. The new rule moved
the start date back to May 1 the year before the Illinois test is given (Illinois
Association of Directors of Title I, 2006). This resulted in the removal of
283,000 students from consideration in determining the AYP levels. As a
result, many children from transient families and low-income communities
were excluded from consideration, and over 50 schools that would otherwise
not have done well were suddenly found to have met their AYP target.

Data Management

School administrators are well educated in the arts of curriculum design,
educational leadership, and the science of budgeting. However, all too often
they are poorly prepared in statistics and data management. This poor level
of knowledge about statistical analysis of test score data is exacerbated by the
timing of the availability of test data.

Often educators do not receive student data reports from the annual
assessment tests until it is too late in the new school year to initiate remedial
efforts. Rather than wait for the state education department's data center to
mail the student and school reports, it is possible to identify children at risk
for failure early. The statistical information that is in the office of any school
can provide an important tool for the early identification of students who will
be at risk of failing.

Within the student files are three important data sources, including the
most recent available assessment test or achievement test scores, school
attendance for the past year, and the scores on cognitive ability tests.[8] Each
of these three data sources is highly correlated with the scores that a child

will earn on upcoming mandated state assessments. Using data from the pre-
vious year, it is possible create a prediction equation for the high-stakes tests
(Creighton, 2007). This is done by using the assessment score actually earned
by former students as one correlation variable (the criterion), and the other
three factors can be combined by multiple correlation analysis into the pre-
dictor. The technology systems of all universities have excellent statistical
software that can be employed for this purpose.

Once the multiple correlation has been calculated by the software, it is
then possible to use the raw score weights associated with each predictor
variable to create a prediction equation. The equation can then be used with
all new student data: cognitive ability, previous assessment scores, and atten-
dance. These scores when weighted by the equation from last year's students
will predict how well each of the new students will do on this year's assess-
ment test.

If all this seems like too much, the rule of thumb is that if a child has any
two of the following three indicators of a potential problem, he or she should
receive extra help preparing for the state assessment test.

Indicators

1. Scores on a test of cognitive ability more than one standard deviation
   below the mean (e.g., IQ is equal to or less than 85)

2. A score on the last assessment test that was below the proficient level,
   or more than one standard deviation below the mean on the stan-
   dardized score report

3. More than 12 days absent from school during the previous 180-day
   academic year

The organization of school data for use in a multiple correlation equation
for prediction can be easily managed by an experienced school technology
consultant. The prediction system can be recalculated and used with the var-
ious subject areas and at all grade levels. A degree of error always exists in
predictions based on correlational data; yet, this model makes it possible for
a school administrator to prioritize students who will need extra services
before they start school in the fall. Each summer, when the data from the pre-
vious year become available, new equations should be established and a new
group of students who are at risk for being below proficient should be iden-
tified. An annually increasing database will also provide for a correlational
prediction model that is ever more stable and useful.

Schedules

Having identified which students are most at risk for failure, it is then possible to initiate instructional efforts designed to improve assessment scores starting at the beginning of the school year.

For such programs to be successful, three things are required from the school's leadership. The first is good public relations. Every effort should be made to make this class a positive experience. It should never become the object of school humor. Second, only the best, most highly effective teacher(s) should be encouraged to teach these classes (Mindish, 2003). Reassignment of such teachers may also involve a trade-off such as extra preparation time or special consideration on the merit pay system. It takes a stellar teacher who is enthusiastic for the task to work with the at-risk students. Finally, the special class should be in addition to the child's regular load. These classes are best taught in a small group setting (N < 16), and they should be scheduled early in the school day.

Time on Task

One misconception about American education is that our children do not spend as much time in school as do the children of other nations. The truth is that our children are taught for about the same amount of time as the median of the world. The nations with the greatest school contact time are Austria and Mexico, each requiring about 1,100 hours, while the median for the various states in the United States is about 1,000 hours (Baker, Fabrega, Galindo, & Mishook, 2004).[9] Research into the relationship between the length of the school day and achievement outcomes has shown that the time factor can only account for about 2.2% of the variance in student achievement. Yet, even the small effect may be critical to schools that have been labeled "in need of improvement" on the basis of test scores. Massachusetts has estimated that it will cost the schools $1,300 per child to add an extra hour of instructional time for a school year (Schemo, 2007). The commonwealth added $6.5 million to the 2007–2008 budget to help pay for a lengthened school day. One plan for using the extra time is to provide the opportunity to reintroduce subjects and curriculum areas that were truncated, or even excised, because of the need for more time to teach high-stakes test skills.

There are two areas where time and achievement may interact significantly. One is the traditional 10-week summer vacation. All children lose ground over the long summer hiatus. The subject areas where this loss is most acute are mathematics and spelling. Perhaps because of summer library programs and

free reading activities, the loss of learned reading ability is not as noticeable with middle-class children as for the children from impoverished communities (Alexander, Entwisle, & Olsen, 2007; Cooper, 2003). Children living in the inner cities experience very noticeable summer loss in reading as well as spelling and mathematics. This may reflect the lack of summer resources available to the families of these children. To combat this academic loss, an increasing number of schools are giving summer reading assignments and mathematics worksheets to complete before the fall term (Bennett & Kalish, 2006).

The best approach to combat the educational loss that occurs during the summer is to institute a program of summer instruction. This can be most effective with elementary and middle-school students, especially those who are at risk for not scoring well on the state-mandated assessments. Under the provisions of the No Child Left Behind Act, it is possible for school districts

Tell me, did you finish that book report yet?

Figure 16.1   Summer Vacation

SOURCE: Cartoon by Merv Magus.

to use Title I funds to establish summer programs for children at risk. These summer programs have even been initiated in the rural farming communities where the children of migrant workers are provided with a stable school environment while their parents are at work (Associated Press, 2005). Control group research conducted with children from the inner city of Baltimore who attended a multiyear summer program revealed that they achieved test scores that were half of a standard deviation higher than those who did not (Borman, 2007). This is a very large gain and may be one way to make a real difference for low-achieving children.

The traditional summer has been from mid-June to the start of September. Recently, a number of school systems have moved the date for the opening of the school year to the beginning of August. By making this move, schools have even more time to cram for the high-stakes tests in the spring. The truncated summer vacation is made up by providing several shorter vacations during the year. This trend has not been missed by either parents or state legislatures. A number of parent groups have sprung up, voicing opposition to the short summer vacations. These politically active groups include Texans for a Traditional School Year and Save Georgia Summers. State legislatures in 11 states have passed laws requiring that public school start classes near or following Labor Day (Bello, 2007).

The amount of time that can be spent teaching reading or mathematics could be increased if the school day were longer. While adding an additional hour each day to the time spent on a subject such as reading or mathematics could improve achievement, the cost would be beyond what our communities could tolerate. The problem is that the extra instructional expense would not be that cost-effective. This reflects the fact that an extra hour of reading instruction each day will only produce an average improvement of one-quarter of a standard deviation on assessment tests measuring reading achievement (Karweit, 1984). A better way to improve scores is to use time more efficiently. As an example, Illinois has a requirement that each elementary level school day include 300 minutes of instruction. An audit found that of the 7 hours each day that children in Chicago spend in elementary schools, only 4 hours are spent on instruction (Smith, 1998).

Time is lost at the start of the school year as well as at end of each school year. Instructional time is lost each day in "homeroom," more is lost at lunch and recess, and more yet for classroom transitions. Instructional days are lost for field trips, assembly programs, fire drills, and pep rallies. The amount of teaching time can be increased by a simple review of the school's schedule and a more efficient plan for the organization of each school day.

One way to change the culture of a school is to think in terms of instructional costs. Consider the cost per minute of any special program that

interrupts the instructional process. If the cost of educating each child is $11,000 per year, and if a principal is planning a 30-minute assembly program for 500 students, the cost of the lost instructional time for that assembly is $2,750. This is found by dividing the cost per child for a year of education by the 1,000 instructional hours in the average school year. This means that each instructional hour costs $11 per child. An assembly of 30 minutes costs $5.50 per child in instructional time.

A large-scale study of the impact of increasing the instructional day for thousands of school children in Boston and Chelsea, Massachusetts, was started in 2006–2007. That study will increase the length of the school day by as much as 3 hours for public school children. It is hoped that the analysis of data from that research will provide a definitive model for the best instructional day for schools to follow (Maxwell, 2006).

CLASSROOM INSTRUCTION

Teachers are at point with the test-based accountability programs now in place in every state. In many ways it is our educators who are the people being left behind. The undergraduate programs they graduated from did not prepare them for the new standards-based accountability assessments, and their administrators have little to offer but encouragement. Yet, it is the classroom teacher who must bring every child up to a level of proficiency in all subjects being assessed. Failure in this task can lead to lower pay (no merit award), whole school penalties, and the loss of respect from all in the community. Parents of a community who have seen good report card grades come home with their children can quickly turn against the teachers, and become major critics of the school, if their children do not pass the state-mandated assessment (Shapira, 2006).

There is an obvious temptation to teach the test and prepare children to answer each of its questions before they must face the measurement. This is not only unethical, it is also illegal. The next obvious option is to "teach to the test." This involves a complete restructuring of the curriculum to stress specific skills assumed to be on the test. This approach to test preparation usually employs a drill-and-fill form of instruction. When a teacher teaches to the test, only those skills needed for the test are emphasized, and inquiry-oriented instruction is eschewed in favor of direct instruction (Firestone, Monfils, Schorr, et al., 2004).

Linda Crocker (2003), past president of the National Council on Measurement in Education, recommends another course. She suggests that teachers "**teach for the test**."[10] Teaching for the test involves knowing how cut

scores are set, understanding how items are developed, and having an in-depth understanding of the content on which the children are to be tested (Boudett, Murnane, City, & Moody, 2005). Teaching for the test also involves the strategic introduction of instructional methodologies that have been shown to optimize the performance of children on the assessments. One aspect of this involves introducing children to the jargon-laden language and logic used on standardized tests. For example, all teachers teach children to find the "main idea" in a paragraph or story being read. Yet, test questions do not ask for this in those words. Tests ask children to "explain what the story is about" or "what is this passage about?" or "what was the author trying to tell you?" Children need to know that the word *passage* means "text" (Greene & Melton, 2007).

Reading

Starting in third grade, all public school children in the United States face a high-stakes test of their reading skill and ability. In 2000, the National Reading Panel (NRP) presented its review and summary of the research on reading, which has been completed and presented since the administration of Lyndon Johnson. From this review of over 100,000 dissertations, presentations, articles, and monographs, a series of conclusions and recommendations emerged (National Institute of Child Health and Human Development, 2000). It is important to note that the work of the NRP was completed before the passage of the No Child Left Behind Act in January of 2002. A number of studies have been published since that time that have examined the relationship between reading instruction and student outcome on state assessment test scores.

The NRP report focused on three levels: parent–child activities, instructional methods in school, and teacher preparation. The major focus for school instruction was on the primary-grade classroom. Here the report presented teachers of kindergarten and first grade with effective strategies for developing and building phonemic awareness. The NRP recommendations included a shift away from phonemic awareness instruction after about a year and the initiation of a full year of well-organized systematic phonics instruction in reading. This effort had the goal of forging the links between the letters of the written language with the sounds of the spoken language. By the end of a year (kindergarten) of phonics instruction, children should have significantly improved their word recognition and pre-spelling skills. During first grade, the teacher's efforts should focus on improving fluency and expanding the child's vocabulary. Before the start of third grade, the NRP recommended the teacher focus reading instruction on text comprehension. One reading program designed to meet the recommendations of the NRP is known as the "Success for All" program. This reading program is widely used

to teach reading to primary-grade children in schools that receive Title I funding under the NCLB Act. For the most part these schools enroll many children from impoverished backgrounds. A well-controlled research study of this kindergarten to third grade reading program has shown that significant improvement in achievement test scores is possible when it is implemented (Borman et al., 2007).

To review the details of the report and the specific recommendations for instructional activities with examples, see www.nationalreadingpanel.org.

In addition to the instructional recommendations of the NRP, there are other steps that teachers of the primary grades can take to improve reading assessment scores for individual children before third grade. The first involves formative monitoring of the development of **prereading** and early reading strategies. This can be accomplished by the use of a reading assessment, which provides measures of basic **benchmark skills in reading**. Once a child's problem area is identified, extra instructional activities can be formulated and introduced to remediate and assist the child develop the missing reading strategy (Good, Simmons, & Kame'enui, 2001; Kame'enui, 2002).

Better assessment scores on state-mandated reading tests are found with the children who are in elementary classrooms that are filled with children's books. This type of environment encourages young readers and expands their reading interests. Classroom instructional strategies for teaching reading that have been shown to correlate with better scores on statewide assessment tests include both prereading and concurrent reading strategies. Prereading strategies are designed to activate the child's background knowledge and help him or her form a purpose for reading. These reading strategies are not intuitive but require direct and explicit teaching (Walkovic, 2003).

Starting at the elementary school level, reading comprehension may be correlated with the background of general knowledge held by the reader. Adults recognize this simple truth when they are asked to read a jargon-laden technical publication. It is easy to read any technical presentation if we have a background in the field. Children who are learning to read are much the same. The falloff in reading comprehension skills after third grade for children with limited experiences and backgrounds is to be expected (Hirsch, 2006). Currently, the normal school practice is to provide students at risk for reading problems with more instruction that is focused on basic skills. The time for this remedial instruction comes from other areas in the curriculum. The loss of field trips and special programs and the truncation of the curriculum are outcomes of the push for better test scores. The irony is that taking time away from general education designed to expand the backgrounds of young readers may be doing harm to their comprehension scores.

Pre-reading strategies that correlate with reading assessment scores among elementary school students include previewing and surveying

(Beserik, 2000). The strategy of self-questioning (forming one's own questions to guide reading) has been shown to improve the reading scores of fifth graders on a statewide assessment (Donnelly, 1999).

Concurrent strategies are those that a child should be taught to do while reading. These are strategies that have been demonstrated to improve reading comprehension scores on statewide assessments. The strategy of picturing what is being read, employing a deliberately slower rate of reading, rereading what was not understood, and outlining or taking notes on what is read are examples of successful concurrent reading strategies (Walkovic, 2003; Wells, 2002). Students who have been systematically taught to employ these reading strategies have significantly better reading comprehension scores on a statewide assessment than those who have not.

At the high school level similar findings were noted with students in the 11th grade (Rex, 2003). Those who took the time to reread sections of passages that they did not understand had significantly better reading comprehension scores compared with those who did not. The nonparametric correlation between self-reports of using this strategy and assessment test scores was quite significant ($r = 0.42$). It is interesting to note that at all three levels—elementary (Beserik, 2000; Donnelly, 1999; Wells, 2002), middle school (Walkovic, 2003), and high school (Rex, 2003)—none of the "post-reading strategies" correlated with assessment scores (Bukowiecki, 2007). Post-reading strategies involve expository writing in response to what has been read. It can also include scanning back over what was read to find the answers to content questions posed by the teacher.

A link has been identified between the writing requirements of high school English teachers and student reading achievement. The more time English teachers spend on analytical writing, and the more homework they assign and grade, the better are the reading test scores (Carbonaro & Gamoran, 2002, 2005). Paradoxically, the amount of time and effort spent in formal instruction of grammar is negatively correlated with reading achievement scores. It seems that the more effective secondary English teachers stress analytical writing, and less effective teachers stress syntax and grammar.

Writing

There are both general and specific strategies that can improve assessment scores that students earn on statewide writing tests. The general strategies that correlate to better writing scores include prewriting strategies such as brainstorming topic ideas with peers, planning the sequence and structure of what is to be written, and formulating a writing plan (Shields, 2000).

The assessment scores in writing are also positively influenced by teachers who provide practice with the writing format used by the state assessment. Teachers who wish to see better assessment scores should be trained in the state's writing test, including the approved rubric and anchors used to evaluate student writing samples (English, 2000). Classroom practice with timed writing assessments that are based on prompts can improve student scores. In such classroom exercises, the teacher should use the state-approved rubric to grade each student paper and explain why that score was assigned.

There is also evidence that the amount of time intermediate-grade elementary school teachers spend on improving student writing skills is directly related to writing assessment scores (Irvin, 2003). The recent addition of scoring software to the education marketplace will facilitate student practice and writing development without a concomitant time investment by classroom teachers. As the states move to the computerized scoring of student writing, there will be an inevitable spread of scoring software in the schools.

Technology provides another way to improve student writing. A few schools have begun to have children produce and write podcasts about their school and communities. These are subscribed to by both children and many of their parents. As this writing activity is very motivating, writing skills have been improving through practice.

A specific strategy that is being employed by teachers to improve student writing is the "five-paragraph essay." By this model, students are taught to read the writing test prompt and formulate their position on the premise that the prompt presents. Then they state their position in a short paragraph. Next, they are taught to write three short paragraphs, each providing an example in support of the position staked out in the first paragraph. The final paragraph is to provide a summation and conclusion. As most writing samples on state assessments only provide students with 20 minutes, students are taught to structure their answers quickly.

Mathematics

There is much less information on the relationship between scores on state-mandated assessments of mathematics and instructional programs than there is on the assessment and teaching of reading and writing. One thing that is well documented is that mathematics scores have changed only marginally over the past few decades (NAEP, 2004). Additionally, the achievement gap in mathematics between minority students and their Anglo-White peers has so far proven to be intractable. Between 5% and 6% of the population of school students have a significant learning problem with mathematics (see

Chapter 13). As with other learning problems, this one may also have a neu-rological basis (University College London, 2007).

In a review of the hundreds of published studies on mathematics instruc-tion by the Institute of Education Sciences (IES) of the U.S. Department of Education, only a tiny handful were found to meet the criteria of a rigorous scientific basis. At the middle-school level, this review found only four pro-grams with a solid scientific foundation, and of those four, only two could consistently demonstrate achievement gains.

To see the full report from the IES, see www.whatworks.ed.gov.

Both of these are computer-based programs. One, Cognitive Tutor Algebra I, provides a full year-long algebra program, including interactive computer-based tutoring, texts, and support material. The second is an online supported instructional program covering the 7th through the 10th grades. As an online program, it is self-paced and allows the student to work through instructional modules from home as well as from school. Further information on these two systems can be found at the following addresses:

www.bcps.org/offices/oit/CognitiveTutor.htm

www.icanlearn.com

Another computer package shown to assist students in passing the math section on a state high school graduation examination is Hotmath.com. This program offers students homework help in mathemat-ics. This is done by providing complete, step-by-step solutions to all of the odd-numbered problems at the end of each chapter of standard texts of mathematics online. The theory behind this is that if a student who is work-ing alone on homework reaches a problem that can't be solved, the com-puter will be able to provide tutoring in the solution to that one or the next one in the series. The help that this online program offers could also be provided by a classroom teacher willing to invest several hours each evening conducting online homework tutorials.

Information about this online approach is available at: www.hotmath.com.

The What Works Clearinghouse of the U.S. Department of Education identified the mathematics instructional program from the Wright Group/McGraw-Hill known as Everyday Mathematics as exhibiting "poten-tially positive" achievement effect (Viadero, 2007b). For more information, see http://ies.ed.gov/ncee/projects/wwc/elementary_math.asp.

One area of mathematics that is particularly difficult for many students involves word problems. These require students to read a problem statement and translate the written words into an equation and then solve for an

unknown. Most teachers teach students to look for words in the problem that describe operations, such as *combined*, which translates to addition. Better results for students on word problem questions and higher scores on statewide assessments have been found when teachers instruct children using a single strategy for problems This method is described as schema-based instruction (SBI) (Jitendra et al., 2007).

   In teaching algebra concepts, Amanda Ross (2007) has demonstrated that there is a latent factor in the instructional practices of the most success-ful teachers. That factor involves the use of constructivist approaches includ-ing inquiry, investigative work, and sharing ideas and explanations among students.

   Mathematics is one subject that is more anxiety inducing for students than others (Cavanaugh, 2007). This anxiety has been shown to interfere with the ability of students to perform well on tests that require arithmetic as well as other more advanced forms of mathematics (Ashcraft, 1995). Evidence is also available that anxiety levels are greatest among children being taught by teachers who have only a sketchy background in mathemat-ics. The confusion experienced by poorly prepared teachers is contagious and translates to heightened mathematics anxiety among the students.

   One area of mathematics instruction that has caught public attention is the use of handheld calculators. The position taken by the National Council of Teachers of Mathematics (NCTM) (2003) is that students need to learn to use paper and pencils to solve problems. Also, NCTM posits that students must learn to estimate what answers should be and should develop the abil-ity to make quick mental calculations. In addition to these core skills, though, handheld calculators should also be part of what goes on in a mathematics class. Tedious calculation of arithmetical expressions by students who know how to do the arithmetic is only a waste of time (Cavanaugh, 2005; NCTM, 1991, 2003, 2006). This is also true for advanced classes with high school students using books of mathematics tables to look up functions and trans-formations that are readily available in advanced calculators.

   A statewide study of the relationship between the use of calculators in middle-school classrooms and mathematics achievement found a direct rela-tionship (Smart Heilshorn, 2003). In this study of the self-reported level of calculator use by 140,000 eighth-grade students in 760 middle schools it was noted that the schools where students scored at the highest levels had students who reported moderate levels of calculator usage. In schools where calculators were used for almost every mathematics class and in those where calculators were rarely used, the mathematics assessment scores were lower.

   Another concern of parents is that heterogeneous grouping has a corro-sive effect on the learning of advanced mathematics by high-achieving

children. In a study of 500 students in middle and senior high school it was demonstrated that heterogeneous grouping improves the learning of advanced mathematics for below-average and average students and has no effect on high-achieving students (Laitsch, 2006).

Test-Prep Curricula

Beginning with the standards-based assessments of the 1990s, there have been a number of educational publishers offering test preparation curricula for the schools. In a study of one such program, an elementary school mathematics curriculum package was found to be significantly less effective than a local curriculum initiative (Kristoback & Wright, 2001). The local curriculum was developed during a summer workshop by the district's elementary teachers working with the state standards and local curriculum experts. The findings suggest that teachers who feel ownership for a curriculum revision are more motivated in their teaching.

ARCHITECTURE AND SCHOOL DESIGN

School Buildings

Each school day one in five Americans spends some time in a school building. Yet, despite this high level of public awareness of schools, the buildings are among the most poorly maintained public buildings. Half of all schools have poor air-quality, containing molds, chemicals, chalk dust, and airborne bacteria. Poor air quality leads to absenteeism, lethargy, and a general malaise.

Our schools are also noisy places where the sound level in most common areas can approach that of a modern airline terminal. Most children do not develop the level of hearing acuity needed to separate a teacher's voice from background sounds until the middle-school years. Competing background noise comes into the classroom from outdoors, surrounding classrooms, mechanical system noises, hallway noises, as well as those generated by two dozen people together in a 12,000-cubic-foot box.

Many of the schools we use today are old enough to have earned historic site designations. The Francis Scott Key Elementary School in South Philadelphia, which serves over 400 children, was built in 1889.[11] Even the new wing on that school was completed 12 years before the First World War (Philadelphia Architects and Buildings Project, 2003). Temporary school facilities (portables) are no longer temporary, with some now approaching

50 years old. Most of the prisons in the United States are cleaner, newer, and in better physical condition than are our public schools. Throughout most of the middle latitudes of the continental United States, public schools are the only public places that are not air-conditioned.

Yet, the total impact of all these conditions adds up to only 2% of the variance among assessment scores. Much of that effect may reflect the fact that our oldest, noisiest buildings with the poorest air quality and the most deferred maintenance are where the children of our poor and minorities go to school.
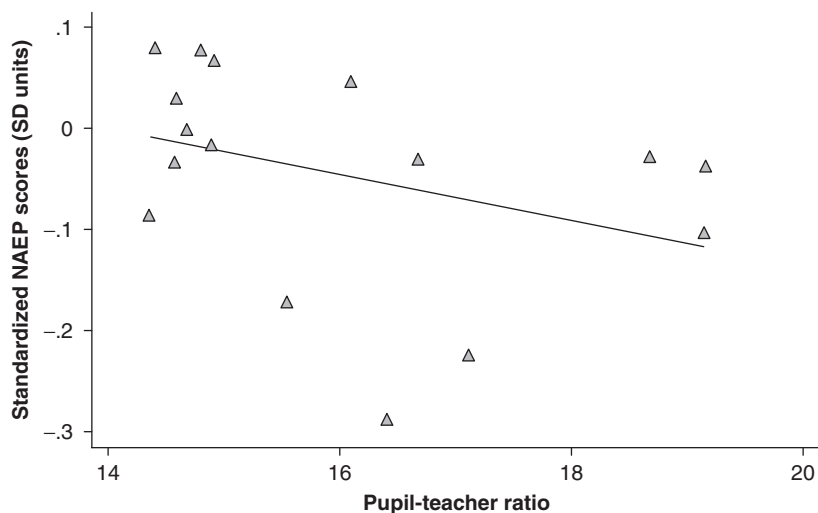
Class Size

Public education is a labor-intensive task. The point at issue always hinges on the question of how many students a teacher will have in his or her class. The 20th-century addition of educational specialists to the schools has made education even less efficient. The requirement that schools provide special education programs for the children that 75 years ago would have been excluded from attending school has also worked to make schools seem less efficient.

School authorities have a vested interest in keeping classes as large as practicable, and teachers know from their personal experience that they are more effective with smaller classes. Research on this question of class size and educational outcomes has reached the consensus that small classes lead to better levels of achievement (see Figure 16.2). The effect of small class size is most pronounced for younger children (K–third grade) and for minority students. In large-scale research, the critical class size is between 15 and 18 students. Each increment over that parameter results in lower achievement. In reducing class size from the mid 20s to 15, the gain in achievement is around 6% (Krueger, 2002, 2003; Word et al., 1990). The statewide research from project STAR in Tennessee found that when African American children are assigned to smaller classes they experience a gain of about 8% on standardized achievement tests. White children assigned to smaller classes experienced only a 3% point gain (Krueger & Whitmore, 2001). This effect is even more pronounced if children are exposed to smaller classes from kindergarten onward (Nye, Hedges, & Konstantopoulos, 2000). It is likely that if African American children attend schools with smaller class sizes from kindergarten onward there would be a 60% reduction in the college admission score gap that now exists between the two groups (Krueger & Whitmore). Most of these positive effects would be greater if teachers were to be provided with specific training in how to work with small classes (Graue, Hatch, Rao, & Oen, 2007). In research involving class-size reduction and

instructional practice, Graue et el. (2007) have shown that teachers typically continue teaching as though they were working with a large group.

A recent review of the literature has shed light on the mechanism by which this improved achievement occurs. Smaller classes actually change teacher instructional practices and improve student engagement. The point can be made that it is more difficult to "loaf" in a classroom if the class size is small than if it is large (N > 30) (Finn, Pannozzo, & Achilles, 2003).

Figure 16.2   Relationship between math and reading NAEP scores and pupil-teacher ratio, 17-year-olds, 1970–96

SOURCE: From "A Response to Eric Hanushek's 'Evidence, Politics, and the Class Size Debate,'" in *The Class Size Debate,* by A. B. Krueger, 2002, Washington, DC: Economic Policy Institute'. Copyright by EPI Publications. Reprinted with permission.

School Size

There is evidence that smaller schools have a positive effect on student graduation rates. In a study of 306 schools systems of rural Arkansas, it was found that small school districts tend to have better graduation rates. There is evidence that small schools or districts have a very small advantage on mandated state achievement tests (Black, 2006). This phenomenon of small schools having better graduation rates is an example of what has been described as the "Hobbit effect." The effect reflects the higher engagement that small schools make possible for students. Co-curricular activities draw a

great proportion of the student body into a school's activities, and classes tend to be smaller in size (Jimerson, 2006).

ADMINISTRATION

Leadership and Outcomes

School leadership at the building level has a clear relationship to the average achievement outcomes for students. The impact of the building's principal can be seen in all aspects of the school. To lead a school to optimal levels of achievement the principal must be a highly motivated and enthusiastic person (Renchler, 1992).

One administrative task linked to the achievement scores of a school is the hiring of new teachers. As was noted in Chapter 15, a new teacher can have an influence on a school over his or her 30-plus-year career. This makes the principal's role in hiring vital. Research by Bruce Baker and Bruce Cooper (2005) has shown that school principals who have strong undergraduate backgrounds tend to recruit, hire, and work to retain teachers from prestigious colleges who have solid academic credentials. This effect was found to be especially true in the schools of the inner cities.

Over the years, a number of studies have demonstrated the link between the leadership style of school administrators and school climate. From the teacher's point of view, school climate can range from a general feeling of malaise and a dread of going to work to a sense of being part of a close-knit community accompanied by feelings of professional fulfillment and pride. School climate can be viewed as the perceived total of all interactions between all the people (adults and children) the teacher deals with. For each teacher that perception of climate sets the tone for how they approach their work and how they view their careers (Goleman, 2006). A negative school climate and the concomitant poor faculty morale can undercut the success of an entire school community. This malaise can be intensified when a school's faculty is sanctioned under the NCLB mandates, thereby creating a cycle of failure and educational depression (Finnigan & Gross, 2007).

The development of a positive school climate takes years to achieve and the commitment of the principal to the goal of a positive school climate. Low morale and a negative climate are much easier to establish, and an ineffective principal can accomplish this in only a few months.

More than any other person, the principal sets the climate of a school. This climate does have a direct impact on the achievement outcome of children (Goleman, 2006). Middle-school research in Texas found successful

school leadership was dynamic, communicative, and interactive (Brown, Claudet, & Olivarez, 2002). Research from California found that school climate as set by the principal was one of three core factors in the relationship between student achievement and leadership. Those things that work to produce good climate include communicating the school goals and the fair enforcement of school rules with students. Also, the principals of highly successful schools worked to open lines of communication both with teachers and between teachers. These school leaders worked to maintain high morale among faculty and were always enthusiastic and optimistic about the learning potential of students (Heck, Larson, & Marcoulides, 1990).

Leadership Style

A meta-analysis of over 60 studies from the United States originally reported during the 1980s and 1990s found that differences in elementary school leadership style could account for a shift of 0.11 standard deviations on measures of student achievement (Witziers, Bosker, & Kruger, 2003). Using data from Pennsylvania, James Cantwell (2007) found the important role that leadership style plays in determining achievement outcomes as measured on a mandated statewide assessment. He first identified elementary and middle-level schools that were in the top quartile in terms of average assessment scores for 3 consecutive years. Likewise, a group of schools in the lowest quartile for 3 years was also identified. The teachers in both groups completed climate questionnaires, and their principals responded to a questionnaire of their leadership styles. One finding from this research was that there is a lack of awareness on the part of the principals of low-performing schools as to what their leadership style really is. The principals of high-performing schools had a clear and accurate perception of their leadership styles. The principals of consistently successful schools are seen by their teaching faculty as more open and approachable, treating teachers as equals. They are seen as up front and easy to understand and are known to look after the welfare of the teachers and staff. These successful administrators required less busy work and routine chores from their teachers. Another difference is that the teachers in the successful schools were more collegial and had better communications with each other and were accepting of their colleagues and sensitive to the problems of other teachers. They were more willing to help new faculty, and they socialized with each other away from school. In a word, they were a community.

The special problem of leadership for inner-city schools was studied in a series of case studies of high-performing schools located in high-poverty communities. This **qualitative research** found that the principals of these anomalous schools had a missionary-like faith in their schools and the

children who attended them. The principals and the faculty shared a team mentality. These administrators would "put it all on the line" by going into classrooms and modeling the teaching of a lesson when one of their teachers had difficulty in making a breakthrough. These highly successful leaders created environments of mutual respect and collaboration with the faculty. They were clever to the point of devious in bringing together a faculty that mirrored their own faith in children and the possibility of success for all children (Kannapel & Clements, 2005).

Summary

The pressure to have students succeed on high-stakes tests can be unbearable. The result is that there is a problem with cheating on the mandated assessments. Beyond out-and-out cheating, there are a number of strategies that parents and educators can employ to optimize both the achievement outcome for children and the apparent success of the school. In examining these strategies, the use of merit pay has not proven to be a method that will improve teacher effectiveness, and it may even trivialize the teaching profession.

Schools can employ statistical models that will prioritize students and identify which students will need extra help preparing for the high-stakes tests. By using the best teachers with the students that are identified, providing extra educational resources to these developmental classes, and maintaining a small class size, these students can have significantly better test scores. Careful planning of each school day to optimize the time-on-task for each student is another administrative activity related to better scores.

A successful school is one where stable leadership has created an atmosphere of trust among all parties. It is a collegial place where the principal and teachers believe in the potential of each child. Such a school is a place where teachers collaborate and share ideas with each other and have a cohesive and mutually supportive team spirit. The principal of such a school is a partner with each teacher in the effort to encourage student achievement.

Discussion Questions

1. Rent and watch one of the films about testing described in this chapter. Write a brief review of that movie. In your review discuss the reality or lack of reality regarding the reactions of adolescent children to the pressures they were experiencing as portrayed in the film.

2. Describe where the "line is drawn" dividing legal from illegal test preparation programs.

3. If you were given the task of developing and offering an inservice program for teams of middle-school teachers on the topic of optimizing student test scores on mandated statewide assessments, what would it include? Provide a content outline for an 18-hour-long inservice activity (3 hours a week for 6 consecutive weeks).

4. Explain and provide examples of the similarities and differences between the ideas of "teaching to the test" and "teaching for the test."

5. What are the characteristics of school leaders whose schools are consistently among the top scoring on high-stakes tests?

Student Study Site

Educational Assessment on the Web

Log on to the Web-based student study site at www.sagepub.com/wrightstudy for additional Web sources and study resources.

NOTES

1. A *juku* is a traditional (private) Japanese cram school attended by 60% of Japan's students after the regular school day. The jukus help students cram for placement examinations.

2. For example, Florida secretary of education John Winn suggested that the old tests be used for practice sessions (Mitchell, 2005).

3. The prize was won by a 15-year-old child who could not drive it or even pay the tax that was due on his winnings.

4. In 1960 more than 90% of the elementary school faculties were women while senior high school faculties were 70% male.

5. In 2004, the national median teacher pay was $47,750. A merit award of 5% would be about $200 per month before taxes. The after-tax net would be about $28 per week.

6. In 1998, Governor Jeb Bush of Florida signed into law a statute requiring school districts set aside 5% of their pay budget for professional teachers and use it to award up to 15% of the teaching staff performance merit pay bonuses. With the

tacit approval of most school systems and minimal cooperation by the teacher's associations, that law was stymied. The state provided no extra money for merit pay. The 5% for merit had to come out of the operating budgets of the school systems. That plan was revised and a new method for merit designed for 2008.

School districts can apply to the U.S. Department of Education for support for their merit pay plans under the Teacher Incentive Fund, which was written into the NCLB Act in 2006 (U.S. Department of Education, 2006)

7. A week after the awards were made the Houston School District realized that it made an error with 100 teachers and asked for the award checks back ("Houston Teachers," 2007).

8. These are variables that have been empirically demonstrated to have high correlations with assessment test scores. Other variables may prove to be more valuable in different school circumstances.

9. Forty-nine states and the District of Columbia require a specific number of teaching days or total hours of classroom instruction. The exception is Minnesota, where the state has no such requirement of the schools.

10. "Some object to regular testing because they believe schools will teach to the test, but if a test measures basic educational skills then teaching to the test means you're teaching a child the basic knowledge of reading and math." President George W. Bush, May 11, 2004.

11. National Historical Register No. 86003296 (12/1/1986). There are several older "one room" schools that are still in use in the rural areas of northern Vermont. One of these was built in 1801 (during the first administration of Thomas Jefferson). In 2007, that school was still used to educate the children of the town of Hancock, Vermont.