

State Accountability Systems

Lance T. Izumi and Williamson M. Evers

For years, proposals for reforming education have centered on increasing inputs, such as more funding for teacher salaries, school construction, and computer equipment. However, with the public disappointed at the lack of major improvement in student performance (despite rising government education budgets), lawmakers have shifted their emphasis to making sure that schools are accountable for the results they produce. As a consequence, many states have created accountability systems that are supposed to ensure that schools in general—and teachers, principals, and district officials in particular—focus on improving student achievement. These accountability programs, however, vary widely both in the means by which they hold schools accountable and in the effectiveness of those means.

State accountability systems are made up of the state academic content standards, the state's tests, and the rewards and sanctions for performance. Whether an accountability

The authors wish to thank Paul Clopton, Jay Greene, Chris Patterson, and Bill Lucia for having read an earlier draft of this paper and for providing suggestions for improvement. They also wish to thank Diallo Dphrepaulezz and Kate Feinstein for research assistance.

system is effective depends on how good these components are and how well they work together.

Margaret Goertz, co-director of the Consortium for Policy Research in Education at the University of Pennsylvania and co-author of a recent report on accountability systems, poses five questions that should be addressed in order to evaluate accountability systems.¹ First, what are schools accountable for? Policymakers must decide what they want to measure, whether it is knowledge of basic skills, attendance and dropout rates, or other indices of performance. Second, for whom are schools accountable? Policymakers must decide whether to hold schools accountable for the performance of all students or to exclude learning-disabled students and non-English-speaking students. Third, does the accountability system measure progress absolutely or relatively? Policymakers must decide whether a set standard must be met (such as a set test score) or whether it is sufficient to show relative progress over time. Fourth, is the accountability system fair? Policymakers need to ensure that all subgroups of students improve their performance. Finally, is the accountability system informative?² For example, testing information must be timely and understandable.

Goertz says that there are three characteristics of successful accountability systems. Such systems encourage schools and districts to have curriculum and teaching that are in line with state standards, to analyze testing data and make use of it in teaching practices, and to make sure students are continuing to make progress year after year.³ According to Goertz, research is showing that school staff want to work

¹“National Accountability Movement Offers Lessons for California,” Palo Alto, CA: EdSource, May 2000: 2–7. This report is an account of the March 2000 EdSource conference on “Ranking California Schools: Will It Improve Student Learning? Undermine Public Support?”

²On the importance of information for accountability, see Caroline M. Hoxby, “Testing Is About Openness and Openness Works,” Hoover Institution Weekly Essays, July 30, 2001, <http://www.hoover.stanford.edu/pubaffairs/we/current/hoxby_0701.html>.

³“National Accountability Movement,” 7.

in “more focused ways” on state and district goals because of accountability systems’ “positive and negative consequences.”⁴

There are other important questions. For instance, if the accountability system is based on state academic content standards, how good are the standards and how aligned are tests with the standards? Also, do the rewards and sanctions really provide an incentive for individual students, teachers, and administrators to improve performance?

The rest of this chapter will apply these various criteria to a sample group of state accountability systems. The chapter looks at California, Texas, and Florida because journalists and policy analysts have said that these states have either comprehensive or innovative accountability programs or both. These three states have accountability programs that are better than those in most other states. Yet the real question that this chapter considers is whether these programs are likely to be effective in improving student achievement.

CALIFORNIA

In April 1999, California passed the Public Schools Accountability Act, the brainchild of newly elected Governor Gray Davis. This accountability law has three major components. First, the Academic Performance Index was to provide individual schools with a numerical score based originally on multiple measures of performance (e.g., test scores, dropout rates, and attendance rates) but which is now, for the time being, based exclusively on student test scores. Second, the rewards program (called the High Performing-Improving Schools Program) would award schools and staff monetary bonuses if they met or surpassed Academic Performance Index growth targets. The third component was the intervention program (called the Immediate Intervention-Underperforming Schools Pro-

⁴Ibid.

gram). The state would intervene in schools that failed to meet targets for improvement in test scores. The intervention-program portion of the legislation also included sanctions (such as state takeover of individual schools) and grants to pay for the interventions.

At present, the state calculates the Academic Performance Index using only scores from the Stanford-9 standardized test. The State Board of Education chose the Stanford-9, which is a multiple-choice, norm-referenced examination, to be the state's assessment device in 1997, several years before passage of the new accountability law. The board chose this test, in part, because as an "off-the-shelf" exam, it was readily available. The state uses the Stanford-9 to test grades 2–11, with students in grades 2–8 tested in reading, mathematics, written expression, and spelling. Students in grades 9–11 are tested in reading, writing, mathematics, science, and history and social science. All students at each grade level take the exact same Stanford-9 exam, and the state requires districts to provide individual student scores to parents. The state Department of Education's Web site publicly lists aggregate Stanford-9 scores by grade level for schools, districts, counties, and the state.

As part of a high-stakes accountability system, such as California's system, the Stanford-9 is less than optimal because it is not aligned to the state's academic content standards.⁵ The state picked out parts of the Stanford-9 that tested the topics listed in the content standards and also developed a set of standards-aligned questions that were added onto the Stanford-9 exam. Later the state expanded this approach into separate standards-based tests (the California Standards Tests). The state gives standards-based tests in reading and mathematics in grades 2–11. Students in

⁵Compare Barbara Miller of EdSource, quoted in Erika Chavez, "High School Test Scores Stagnant: Frustrated Educators Are Trying to Figure Out Why Teens Aren't Doing Better," *Sacramento Bee*, August 17, 2001.

grades 4 and 7 are tested on writing, and those in grades 9–11 are tested on the state content standards in science and in history and social science.⁶ Students' performance on these standards-based questions have not yet become part of the Performance Index calculation, although they will be included in coming years.

In the future, California plans to retain the Stanford-9 (or an equivalent nationally normed test) both as a “basic skills” test and as a check on how California's self-assessed progress measures up against national norms of performance.⁷ But in 2002, the state will shift emphasis away from the Stanford-9 and will rely mainly on its standards-based tests for accountability. These tests reflect the state's 1997–98 academic content standards, which California educational officials, California school teachers, and national experts alike consider to be quite high, probably the highest in the country.⁸

But thus far the state has exclusively used Stanford-9 for accountability. In addition to this test not being based on the state standards, there is another drawback to using it as a tool of accountability. Stanford-9 questions do not change from year to year.⁹ As the state Legislative Analyst's Office, the nonpartisan research arm of the State Legislature, points out, “Particularly with a high stakes test, it is important to vary test questions from year to year in order to minimize possibilities for literal ‘teaching

⁶In late 2001, the State Board of Education announced plans for a history test in grade 8 (and elimination of the current history test in grade 9) as well as a science test in grade 5.

⁷Suzanne Tachney in David Fleishhacker and Tacheny, “Arguments: Do Tests Add Up?” *San Francisco Chronicle*, September 2, 2001; Larry Crabbe, quoted in Erika Chavez, “School Test Results Out This Week,” *Sacramento Bee*, August 13, 2001.

⁸Tachney in Fleishhacker and Tacheny; Erika Chavez, “Kids Fare Poorly on New Test: Just Three in 10 Meet the State's Standards in Language Arts,” *Sacramento Bee*, August 16, 2001.

⁹In contrast, the California Standards Tests are “refreshed” with a significant proportion of new questions each year.

to the test' and outright cheating."¹⁰ With increasing frequency, schools have used bootlegged copies of the test, and officials have had to discipline school personnel for cheating.

When questions do not change from year to year, test scores tend to rise each year as students become more familiar with the sort of questions on the test. Increasing scores may not, therefore, represent increases in true learning but rather test preparation and gaming strategies. According to Joan Herman, co-director of the Center for Research on Evaluation, Standards, and Student Testing at UCLA, scores usually drop when test writers introduce new questions or forms and then rise again as the test becomes more familiar.¹¹ Says Herman, testing specialists concur that "it's not a good idea to give the same test form from year to year, or use exactly the same test items."¹² "By changing test forms or changing the items," says Herman, "you prevent schools from over-focusing on the specific items that are on the test."¹³

For 2000–2001, though, the Academic Performance Index used only the Stanford-9 scores and did not use any scores from the augmented standards-aligned tests. Based on the Stanford-9 results, the state Department of Education calculates a score ranging from a low of 200 to a high of 1,000 for each school. The interim statewide Academic Performance Index target for all schools is 800. The state Department of Education also ranked schools on a one-to-ten decile ranking scale, with ten being the best. The department uses a separate "similar schools" ranking to compare schools with other schools having similar demographic characteristics.

¹⁰"Analysis of the 2001–02 Budget Bill," Sacramento, CA: Legislative Analyst's Office, February 2001: E-93.

¹¹"California's Student Testing System: Hard Choices and New Direction," Palo Alto, CA: EdSource, June 2001: 7. This report is an account of the April 2001 EdSource forum on "Tests and More Tests: The Road Ahead for Student Assessment."

¹²Ibid.

¹³Ibid.

Because the Stanford-9 is a norm-referenced test, the state reports results also in terms of national percentiles. The state uses three performance levels or cut-points, the twenty-fifth, fiftieth, and seventy-fifth national percentiles, to create school, district, county, and state scores. The state does not use any further performance levels or criteria in reporting Stanford-9 results.

Schools scoring below 800 must close the gap between their current score and the state performance target by at least 5 percent to meet their growth target for the year. For example, if a school's 1999 Academic Performance Index score was 500, the school's growth target would be $(800 - 500) \times 5$ percent = 15 points.¹⁴ Yet with these growth targets, the state requires so little improvement that, for many schools, meeting the state's target score of 800 would take decades.

Each numerically significant ethnic or socioeconomically disadvantaged subgroup at a school (that constitutes at least 15 percent of the school's total pupil population and consists of at least thirty students) must have a growth target of 80 percent of the school's growth target. Thus, if a school's growth target was 15 points, each numerically significant subgroup at the school must improve by at least 80 percent of 15 points, that is, by 12 points.

The rewards portion of the accountability system includes several programs that are triggered when schools meet their growth targets and subgroup targets and have 95 percent Stanford-9 participation in grades K–8 and 90 percent participation in grades 9–11. One program, the Governor's Performance Awards, sends state grants to individual school-site councils, which have the discretion to use the funds as they see fit. In addition, state grants from the School Site Employee Performance Bonus Program are to be divided equally among school site councils and all school site staff. Finally, the Certificated Staff Performance Incentives Program targets staff at low-performing schools that have the

¹⁴The State Department of Education uses this example.

highest growth rates. The program awards \$25,000 bonuses to 1,000 staff statewide, \$10,000 bonuses to 3,750 staff members, and \$5,000 bonuses to 7,500 staff members.

The intervention program originally applied to those schools that scored below the fiftieth percentile on the Stanford-9. More than 3,100 schools fell into this category. The state changed the program in 2000. Now schools that rank in the lower half of the Academic Performance Index and fail to meet their growth targets are eligible to apply to receive state interventionary assistance. Under these requirements, 938 schools were eligible. Once eligible, however, not all low-performing schools become part of the intervention program. Participation is voluntary, and schools must apply. Once schools apply, state officials select only 430 schools each year. As a result, many low-performing schools neither apply nor are selected for the program. For example, in 2000–2001, of the 938 eligible schools, only 532 applied for the 430 slots. In other words, 406 eligible low-scoring schools voluntarily decided not to apply, and of those that did, 102 were not selected.

The upshot is that some of the worst schools in the state, such as those ranked at the 1 level, are neither compelled to apply for the program nor guaranteed selection if they do apply. It is quite possible for a school with a 4 ranking to be chosen over a school with a 1 ranking. Further, the school with the lowest Academic Performance Index score in the state may not be eligible for the program if it meets its incremental annual growth target.¹⁵ Thus, such a school would be ineligible to apply to the program, whereas a higher scoring school may be eligible if the latter failed to meet its growth target.

Once selected for the program, schools in the first year receive \$50,000 in planning grants to develop a comprehensive school reform plan. As part of the planning phase,

¹⁵Legislative Analyst's Office, *op. cit.*, E-101.

schools must hire qualified external evaluators to assist in developing the reform plans. The plan must then be approved by the state Board of Education. After approval, the school receives annual implementation grants of up to \$200 per enrolled student. Schools receive the implementation grant for two years and may be granted a third year of funding if they continue to struggle to meet their Academic Performance Index growth targets.

A major weakness of the reform plan process is the lack of quality control over the recommendations made by the external evaluators. Some schools have complained that the external evaluators do little except put the ball in the court of teachers and principals to come up with a reform plan. Proponents of solid education have complained that all too many evaluators recommend progressive student-centered teaching methods—such as discovery learning—that, according to the evidence, do not improve student achievement.

Schools in the program that fail to meet their growth targets may be subject to a number of sanctions. If a school fails to meet its target but there is evidence of significant growth, the school can continue in the intervention program and continue to receive funding. If the school fails to meet its growth target in twelve months and doesn't show significant growth, state officials may reassign the staff, negotiate various site-specific changes, or try other approaches. If a school fails to meet growth targets and to show significant growth after twenty-four months, the state superintendent of public instruction may take over the school, reassign the principal, and, in addition, do one of the following:

- Allow students to attend other public schools
- Transform the school into a charter school
- Turn the management of the school over to another educational institution
- Reassign the teachers
- Negotiate a new labor contract

- Reorganize the school
- Close the school

In fall 2001, Governor Davis signed a bill that layers another state funding program on top of the existing intervention program. Under this funding program, low-performing schools could receive a grant of \$400 per student—double the previous amount granted for implementation. However, instead of being subject to independent external evaluation, as is the case under the present intervention program, the new funding program allows districts to evaluate their own schools. Further, the sanctions timeline would be lengthened. Rather than face sanctions after two years, schools that fail to improve get another year to raise their performance (and this time only under district supervision). Even if they do not hit their annual testing growth target, schools can then get another year's reprieve if they simply show "significant" growth. It may be four or more years before any sanctions are imposed on failing schools.

Every year, local school boards must issue a School Accountability Report Card for each school within the district. Each report card must include the most recent three years of testing data for student achievement by grade level in reading, writing, arithmetic, and other academic goals. Secondary school report cards will also list the percentage of seniors taking the SAT college admissions test and their average grade on the test. Also reported are dropout rates and suspension and expulsion rates, plus progress toward reducing class size. Finally, the report cards list the number of days of staff development and the number of credentialed teachers, emergency credentialed teachers, teachers without credentials, and teachers working outside their areas of competence. The local board must send these accountability report cards to all parents.

Although not formally part of the 1999 accountability law, California has also created a High School Exit Exam (HSEE) aligned to the state's academic content standards.

The American Institutes for Research developed the test for the state, and it covers English-language arts through the tenth grade and math through Algebra I. The exam is graded on a pass-fail basis, and students in the class of 2004 must pass the exam to receive their high school diploma. In March 2001, about 350,000 high school freshmen took the HSEE for the first time. A state advisory committee recommended that the cut-point for a passing score be 70 percent. However, because of the poor performance by the students tested, the state Board of Education reduced the cut-point for passing to 55 percent. Students will also get to take the HSEE multiple times if they fail to achieve a passing score.

Ever present in the background of the accountability law are California's academic content standards. The state's standards are among the best in the nation. In its 2000 survey of state standards, the Thomas B. Fordham Foundation ranked California's standards for English, mathematics, history, and science as number one in the country or tied for number one.¹⁶ California's standards are rigorous, specific, comprehensive, and cumulative.

Given this description of California's accountability system, what can we say about how well the system addresses the Goertz criteria set forth at the beginning of this chapter? First, what are schools accountable for? Under California's system, this has so far been unclear. Because the Stanford-9 is not aligned with the state academic content standards, it is uncertain whether schools should be accountable for the material tested on Stanford-9 or the content contained in the standards. Also, when the California legislature enacted the state's accountability law in 1999, it said that the Academic Performance Index was to include such nontest measures as graduation rates and student and teacher attendance rates. For the time being,

¹⁶See "The State of State Standards 2000," Chester Finn and Michael J. Petrilli, editors, The Thomas B. Fordham Foundation, Washington, DC, January 2000.

though, the state has excluded these nontest measures when it calculates the index. Governor Davis has, in fact, turned down recommendations from the state superintendent of public instruction to include them. The Davis administration is supposed to conduct a study that will help set a timetable for including these measures in the Academic Performance Index calculation. Such uncertainty makes it difficult for schools to determine if they will be held accountable for these measures or not.

For whom are schools accountable? The California system is clear on this point. All students, with the exception of a small number of students with individual education plans that exempt them from testing and students with parents or guardians who request an exemption, must take the Stanford-9, including students who do not speak English. Spanish-speaking students who do not speak English and have been in the California school system for less than twelve months must take both the Stanford-9 and an additional test in Spanish. Some renegade school districts in California, however, such as San Francisco Unified, have refused to test large numbers of non-English-speaking students. In those districts, it is less clear for whom the schools are accountable.

Is progress measured absolutely or relatively? In California, the state government measures the progress of schools relatively through annual growth targets for test scores. As mentioned earlier, little growth is called for, and it will take years for many schools to reach the state-recommended Academic Performance Index performance target of 800 even if they meet their annual growth goals.

Is the accountability system fair? Although California's accountability system is based to a large extent on aggregate test scores, the state also requires that various subgroups, including ethnic minorities and the socioeconomically disadvantaged, improve their performance. With this requirement, it is less likely that improvement efforts will ignore some group of students.

However, because California's rewards and sanctions program is voluntary, children are trapped in poor-performing schools that refuse to participate in the state program. Such a situation is not fair to the students in these failing schools and to the schools that recognize their deficiencies and actually try to improve.

Is the accountability system informative? The answer is yes and no. Local districts report individual test scores to parents. However, for the general public, the state reports scores and rankings on a school-by-school basis using schoolwide averages. No one reports individual student scores longitudinally, so it is impossible to tell which teachers are effective and which are ineffective in improving student achievement. Also, many school districts are not able to track student achievement from one year to the next, with the result that a teacher may have no prior-year testing information about students. Further, even in districts with well-integrated information systems, the district may receive no testing information for students coming from another district. Now the state is in the process of establishing an Academic Performance Index database that will include longitudinal data on individual student test scores. This database could be used in the future for a value-added analysis showing how much individual students improve their performance each year. This database could also allow districts to provide information on test results to teachers and (in the case of transfer students) to other districts.¹⁷

In addition to the Goertz criteria, there are several other key questions. For instance, under California's accountability system, how much are assessment and classroom instruction aligned with the state standards? As noted, although California has an extraordinarily good set of standards, there have been major alignment problems with the accountability system. The state Legislative

¹⁷See Legislative Analyst's Office, *op. cit.*, E-127.

Analyst's Office aptly points out the dilemma created by nonalignment:

Since the Stanford-9 tests a different set of information than the content standards, the state is sending conflicting messages about what schools should be teaching. Should teachers and administrators focus instruction on areas covered by the Stanford-9? The state is offering bonuses that can exceed \$25,000 per individual for success measured by that test. Or should they teach what is expected under the academic content standards for which the state has invested \$1 billion for new textbooks, the Governor is proposing \$335 million for staff development, and upon which the [High School Exit Exam] is based?¹⁸

California, however, is in the process of making a transition to standards-aligned assessment. The state Legislative Analyst's Office says that, assuming there are no problems with the validity of the grading standards or the tests themselves, it would be feasible to include results from the standards-based tests in the Academic Performance Index for 2002.¹⁹ Once a standards-aligned assessment is officially part of the accountability system, there will be greater incentive to teach to the standards as opposed to teaching to a nonstandards-aligned test (the Stanford-9). Given the high quality of California's standards, such a development should help improve learning in the classroom and increase true student achievement. The state also will be trying to help teachers by spending a significant amount of money on training already-hired teachers to teach the material in the state standards.

Finally, do the rewards and sanctions in California's accountability system give students, teachers, and administrators real incentives to improve? For some schools perhaps, but for many others the answer is no. The incremental growth targets for low-performing schools are quite small.²⁰ Also, participation in the intervention program is voluntary,

¹⁸Legislative Analyst's Office, op., E-92.

¹⁹Ibid., E-96.

²⁰For schools scoring above the state-recommended API target of 800, the annual growth target is a single point. Ibid., E-94.

and even those that apply may be turned down in the state selection process. Thus, the lowest performing schools in the state may not be participating in the intervention program and therefore remain beyond the reach of the state's improvement efforts and sanctions. Further, if the lowest performing school in the state meets its minimal growth target, it is ineligible to apply for assistance under this program. Because the intervention program is voluntary and selective, out of the more than 3,100 schools in California that fall below the fiftieth percentile on the Stanford-9, only 830 are participating in the program.

There are other significant omissions from the California system that dilute the incentive to improve. Unlike Florida, students attending a failing California school are not eligible for a state-funded exit voucher to attend a private school. This lack of a parental choice mechanism can trap students, especially students from poor households, in a failing public school. Without having to worry about the potential loss of their customer base, public school officials have less incentive to improve student achievement.

Also, under the 1999 accountability law, there is little incentive for students to do their best on the accountability tests. Sanctions do not fall on them if they fail to perform their best on the tests, although a bill passed in 2000 does allot increased funds for students who do well on the standards test. Some commentators say that high school student scores on the tests are as low as they are because high school students give only a token effort when taking the tests, since they suffer no repercussions for such half-hearted behavior. This is not the case with the HSEE, where students may not receive their high school diplomas if they fail to pass the exam.

As for teachers, the state rewards them with bonuses if they work at schools where students perform exceptionally better than before. However, in the accountability system, rewards and sanctions are not tied to individual teacher performance. Poor teaching by an individual teacher does not automatically invite any sanction. Yet recent research shows that teacher

quality and effectiveness is perhaps the most important factor in determining student achievement. The lack of a nexus in the California accountability system between individual teacher performance and student performance on state assessments constitutes a huge hole in the accountability system.

It should also be noted that the state's schools of education, which prepare and produce California's teachers, are also absent from the accountability system. If too many teachers in California are not performing well, the state's schools of education must bear some of the responsibility. Yet they are left completely out of the accountability equation.

Ultimately, it is up to the discretion of the elected state superintendent of public instruction to use many sanctions available under California's accountability system. Political concerns clearly play a role in this process. Over the last several decades, the state teachers unions have been strong allies of the state superintendents. Given the unions' antipathy to charter schools and renegotiation of collective bargaining contracts, it is difficult to imagine a union-allied superintendent using many of the discretionary powers allocated under the 1999 accountability law. Thus, many sanctions may end up as mere paper tigers, depending on who is occupying the superintendent's office. Such a situation, again, argues in favor of a sanction, such as Florida's exit-voucher option, which places sanction power in the hands of parents rather than a politician.

As of fall 2001, the state had not imposed sanctions on any school in California because low-performing schools in the accountability program were in the second year of the two-year period in which they had to show improvement.²¹ The State Legislature was still in the process of determining what types of warnings the state would give to nonimproving low-performing schools and when it would deliver those warnings.

Under the California system, principals, superintendents, and school boards are held to some accountability. Princi-

²¹The authors benefited from discussions with Roger Magyar on the material in this paragraph.

pals may be reassigned, and the state can take over schools. However, if no one holds teachers individually accountable for their effectiveness, is it fair to hold principals and school boards accountable? Is it fair to hold a principal accountable for student performance at his or her school if he or she cannot hold individual teachers at the school accountable for their performance?

The sanctions portion of the California system, thus, is porous and inadequate. Add on the other deficiencies in the system and we must conclude that not only is accountability in California a work in progress, but it also still needs work.

TEXAS

Texas' business leaders were worried about the quality of the state's workforce and wanted more accountability in the schools. They worked with politicians and parents to promote school reform. The state began by creating its own tests and then started to rate schools based on their test scores. This rating and accountability system has led to gains in student learning.

In 1993, the Texas Legislature passed statutes that mandated the creation of the Texas public school accountability system to accredit school districts and rate schools. The foundation of the Texas accountability system has been the state test, the Texas Assessment of Academic Skills (TAAS), first implemented in 1990. The major parts of the accountability system, including student accountability, school accountability, the performance database, the accountability rating system, and the rewards and sanctions program, are all based to a great extent on TAAS results.

The TAAS exam tests reading and math in grades 3–8, writing in grades 4 and 8, and science and social studies in grade 8. Also, there are exit examinations in reading, mathematics, and writing in grade 10 and end-of-course examinations in English, U.S. history, biology, and Algebra I. Whereas

California has used the norm-referenced Stanford-9, TAAS is not a norm-referenced test. It is a criterion-referenced exam that is designed to measure competency on Texas' statewide curriculum. The curriculum was first implemented in 1985 and later updated by the Texas Board of Education in 1997. The state department of education has produced "educator guides" that are designed to show what components of the curriculum may be tested on the TAAS.

TAAS tests all students, rather than a sample, and uses largely multiple-choice questions. The TAAS writing assessments in grades 4, 8, and 10 include short-answer questions and essays.

It is widely believed that students must answer correctly 70 percent of all items on a TAAS test in order to pass. The reality, however, is much more complicated. Although the Texas Board of Education in 1990 adopted a 70 percent standard for passing/minimum expectations in writing, reading, and mathematics, the state department of education says that, "The passing standards for the TAAS and the end-of-course tests are related to two factors: 1) the difficulty of the items on the tests and 2) the number of the items students have to answer correctly to pass."²² Because test items fluctuate in difficulty from year to year, the state department explains that the performance standard for passing is adjusted:

For instance, suppose a test contains fairly easy items when a standard is set at 70%. A subsequent test is administered with slightly more difficult items. If the standard of 70% of the items on the test were used exclusively, the students taking the second test would be held to a higher standard than the students taking the first test. The percent of items required to pass would be the same, but the difficulty of the items would not be. In order to set the standard on the second test to an achievement level equivalent to that of the first test, the tests are equated, and the percent required to pass is adjusted. In this case, the percent of

²²"Texas Student Assessment Program Technical Digest of the Academic Year 1999-2000," Texas Education Agency, NCS Pearson, Harcourt Educational Measurement Inc., and BETA, Inc., 2001: 36.

the items required to pass the second test would be less than 70%, since the items were more difficult.²³

Because Texas is implementing a new set of academic content standards, the more rigorous Texas Essential Knowledge and Skills (TEKS), more difficult questions have been added to the TAAS exams in recent years. Thus, in an extreme example of passing-score adjustment, on the 2001 sixth grade TAAS math test, a student had to answer correctly only twenty-eight out of fifty-six questions, or only 50 percent, in order to pass the exam. In 1998 and 1999, a sixth grader would have had to answer correctly thirty-eight out of fifty-six questions in order to pass.²⁴

Texas education officials defend the scoring adjustment process by saying that it is needed in order to compare current score results with results from previous years. However, there is a problem with this logic: If students in Texas are now expected to master the material in a new set of academic content standards (namely, TEKS) that is supposed to be more rigorous, then state education officials must believe that students should be performing at a higher level than was previously expected of them. To adjust passing scores downward in order to compensate for the increased difficulty of the academic content standards contradicts the purpose of the new standards, which is to raise the level of student performance.

Further, because the percentage of students passing the TAAS, 82 percent, rose again in 2001, the fourth increase in a row, testing experts cast doubt on the scoring adjustment process. Walter Haney of the Center for the Study of Testing, Evaluation, and Education Policy at Boston College says, "The lower minimum scores to pass [math] could be a factor in the increased percentage of students who passed the TAAS this year."²⁵ Haney also noted that passing rates

²³Ibid.

²⁴"Bar for Passing TAAS Lowered," *Dallas Morning News*, June 9, 2001, available at <http://www.dallasnews.com/cgi-bin/print/cgi?story=.../389136_taastranspassing_09.htm>.

²⁵Ibid.

typically drop, not increase, when tougher test items are introduced.²⁶

The Texas accountability system has several key components. First, there is the school report card that all schools must compile and give to each parent. The state department of education (called the Texas Education Agency) created the requirements for the school report cards. Cards must include a school's TAAS passing rates by ethnic and socioeconomic subgroup; the average TAAS passing rates for state, district, and school; and test-taking exemptions for regular, non-English-speaking, and learning-disabled students by subject for ethnic and socioeconomic subgroups. In addition, the school report card must include information on attendance and dropout rates, end-of-course exam participation, the student-teacher ratio, the completion rate for the Recommended High School Program, and administrative and instructional costs per pupil.

Although the school report cards are supposed to put pressure on schools to do well by keeping parents informed, the Texas system also contains more concrete measures to hold students and schools accountable. Students must pass the four high school end-of-course exams or the exit exam in order to graduate. Schools are subject to various rewards and sanctions (which will be discussed later in this section) based on TAAS scores.

Texas has also devised an accountability rating system that uses TAAS performance (without the end-of-course exam results), dropout rate, and attendance rate to calculate annual progress for schools and districts. The state has rated schools and districts since 1994. The rating system has four performance levels:²⁷

²⁶Ibid.

²⁷Margaret E. Goertz, Mark C. Duffy, and Kerstin Carlson-Le Floch, "Assessment and Accountability Systems in the 50 States: 1999–2000—Assessment and Accountability Systems: 50 State Profiles," Consortium for Public Policy Research in Education, University of Pennsylvania Graduate School of Education, March 2001 (CPRE Report Series RR-046), <http://www.gse.upenn.edu/cpre/Publications/Publications_Accountability.htm: TX 3-4>.

1. Exemplary (90 percent of total students and each ethnic and socioeconomic subgroup passing each subject, a dropout rate of 1 percent or less for all students and subgroups, and an attendance rate of 94 percent or higher)

2. Recognized (at least 80 percent of total students and each subgroup passing each subject, a dropout rate of 3.5 percent or less for all students and subgroups, and an attendance rate of 94 percent or higher)

3. Academically Acceptable/Acceptable (at least 50 percent of total students and each subgroup passing each subject, a dropout rate of 6 percent or less for all students and subgroups, and an attendance rate of 94 percent or higher)

4. Unacceptable/Low-Performing (below 50 percent of total students and each subgroup passing each subject, a dropout rate of above 6 percent for all students and subgroups, and an attendance rate below 94 percent)

The state determines that a school or a district has made adequate yearly progress if it achieves an “acceptable” rating.²⁸ In addition, a district cannot be rated Exemplary or Recognized if it has one or more Low-Performing schools or has 1,000 or more, or 10 percent or more, students in grades 7–12 who were unreported on either the comprehensive student-level information system enrollment record or the school dropout record.

In order to calculate the information necessary for the accountability rating system and for other elements of the general accountability system, Texas combines demographic information and performance data in its Academic Excellence Indicator System. The Indicator System brings together:

1. TAAS passing rate
 - by grade
 - by subject

²⁸Ibid., TX 4.

- by all grades
 - by subgroup
2. End-of-course exam passing rates
 3. Annual attendance rates
 4. Annual dropout rates
 5. High school graduation rates
 6. Percent of high school students completing an advanced course
 7. Percent of high school graduates completing the Recommended High School Program
 8. Advanced Placement and International Baccalaureate results
 9. SAT and ACT participation and results²⁹

Like California's Similar Schools Index, Texas uses a statistical tool (called Comparable Improvement) to compare schools with similar characteristics and to measure school-level growth. Under Comparative Improvement, a school is compared with a forty-school group that shares similar ethnic, socioeconomic, student mobility, and non-English-speaking characteristics. Comparative Improvement is based on the Texas Learning Index, which uses TAAS reading and math in grades 3–8 and in exit exams. Texas Learning Index growth scores are calculated using student Texas Learning Index growth (an individual student's current-year score in math and reading minus the student's prior-year score) and campus average Texas Learning Index growth (sum of student Texas Learning Index growth by subject divided by the total number of students by subject). Schools are ranked in quartiles based on Texas Learning Index growth. Also calculated are the number of students at each school meeting or exceeding the Texas Learning Index growth standard and the number of

²⁹Ibid., TX 3.

students performing at high or low levels. The state distributes a Comparative Improvement report with each school's Indicator System report.

Once the state makes all these various ratings and calculations, the state bestows a variety of rewards and sanctions on schools. If a school has an Exemplary, Recognized, or even Acceptable rating, it is eligible to participate in a rewards program called the Texas Successful Schools Award System. The state awards school bonuses of between \$500 and \$5,000 if a school's Comparative Improvement quartile ranking, based on Texas Learning Index average growth, puts the school in the top quartile of the forty-school comparison group in reading and math.

At the other end of the spectrum, a Low-Performing school is subject to a number of interventions and sanctions, including having the local school board hold a public hearing, submitting an improvement plan to the state, holding a state hearing, appointing a special on-site intervention team, appointing a board of managers from district residents, or closing the school down. If the state appoints a special intervention team, the team evaluates the school to determine the cause of poor performance, recommends remedies, assists in developing a school-improvement plan, and helps monitor the school.

Few schools, however, ever receive the harsher sanctions available to state officials.³⁰ For example, the state commissioner of education can send monitors, with limited authority, into schools to make suggestions of remedies to the local school superintendent. If a school makes inadequate progress after the appointment of a monitor, the state sends in a master, who has full authority over the local school board and superintendent. Currently, there are seven schools in Texas under the supervision of a monitor and two under the supervision of a master.

³⁰The authors have benefited from discussions with Chris Patterson on this topic.

In addition, although the state may reconstitute a failing school (an entire staff may be transferred and a new one brought in) or close it down, only a handful of schools have ever been subject to these sanctions. Indeed, the state has reconstituted only three schools, all in the San Antonio Independent School District, under the Texas accountability program. The state commissioner may also put failing schools under an outside management board. Such an action, considered a severe penalty, has only occurred once in Texas.

For low-performing districts, the state education commissioner may also conduct a Special Accreditation Investigation to examine situations identified through complaints, low performance ratings, or state-initiated analyses. Specifically, an accreditation investigation can be initiated if there are severe problems in governance, finances, testing practices, data quality, education of the learning-disabled, compliance with federal regulation, or administrative management. If the results of an accreditation investigation indicate that the state should impose sanctions, the commissioner may take away the district's accreditation, lower the district's accountability rating, or both.

Texas does not hand out special grants or additional resources to Low-Performing schools, although state regional education service centers can contact a district about various services available to such schools. Assistance to Low-Performing schools includes data analysis, identification of problems, information about effective practices, and curriculum alignment.

The accountability system not only provides information to parents about which schools are performing poorly, but it also provides some relief to parents whose children are attending Low-Performing schools. As part of the accountability program, in 1995, the Texas legislature created the Public Education Grant Program that permits parents with children attending a Low-Performing school to transfer to another public school, even one outside district boundaries, that had higher performance results.

Under the Texas accountability system, what are schools accountable for? The simple answer is: The accountability rating system makes schools accountable for test scores, dropout rates, and attendance rates. However, because so much of the state's accountability system relies on the TAAS, the real and more complex question is: What does the TAAS hold schools accountable for? It is this question that has stirred up a hornet's nest of controversy in Texas.

Over the years, the passing rate on the TAAS has increased substantially. In 2000, the percentage of students passing the reading, math, and writing sections of the TAAS exams in grades 3–8 and grade 10 was 80 percent, up from 56 percent in 1994. Poor and minority students made the highest gains. Passing all sections of the tenth grade TAAS exit exam is a requirement for graduation. In 2000, 80 percent of tenth grade students taking the tenth grade exam passed all sections, up from 50 percent in 1994. If they fail to pass the tenth grade TAAS, students have eight chances to pass it again. Among students who took the exit exam in 1998 and who were in the same school district in 2000, 92 percent had passed the test.³¹ For all these gains, though, the question is whether these increases in passing rates represent real increases in learning.

For those who argue that TAAS score increases do not represent improved learning, the biggest complaint against the TAAS exam centers on its lack of rigor. Jeff Judson, president of the Texas Public Policy Foundation, notes that the TAAS is a minimum proficiency test, not an achievement test, and, therefore, "A perfect score means the student has met minimum expectations."³² Judson observes that because a student

³¹Craig D. Jerald, "Real Results, Remaining Challenges: The Story of Texas Education Reform," *The Business Roundtable and The Education Trust*, April 2001: 11–12.

³²Jeff Judson, "A Review of the Texas Public School Accountability System," *Veritas*, Winter 2001: 34.

can get 30 percent of TAAS questions wrong and still pass the exam, “students are passing the test even though they are only able to answer questions that are one, two, or three years below grade level.”³³ Indeed, a team from Mathematically Correct found that from 1995 to 1998, questions on the fifth grade TAAS exam, for example, could be written at the third, fourth, or fifth grade level (as outlined in Texas’ state standards). Questions, therefore, averaged approximately one year below the grade level being tested.³⁴ Many observers say that, at best, the tenth grade TAAS exit exams measure only eighth grade-level knowledge and skills.³⁵ One analysis found that achievement at the sixth grade level, as measured by the California math standards, was sufficient to pass the TAAS exit exam.³⁶

In its recent analysis of TAAS scores, the Dallas school district found that passing the TAAS reading exam in grades 3–8 was equivalent to performing only at the twenty-fifth percentile on the Iowa Test of Basic Skills (ITBS). Passing the TAAS math exam in grades 3–5 was equivalent to the fortieth percentile on the ITBS, whereas passing the TAAS math exam in grades 6–8 was equivalent to the thirty-third percentile on ITBS.³⁷

Others defend the TAAS exams, saying that the rising TAAS scores are matched by improved student performance on other tests, such as the National Assessment of Educational Progress (NAEP). For example, on the 1996 NAEP fourth grade math exam, Texas ranked among the top states, and its students tied for the greatest increase in scores from 1992 to 1996. Also on that test, Texas African American and Latino students outscored students from the same groups in

³³Ibid.

³⁴Paul Clopton, “Texas Mathematics Education in Transition,” *Texas Education Review*, Fall 2000: 57.

³⁵Craig D. Jerald, op. cit., 7.

³⁶Paul Clopton, op. cit., 57.

³⁷Dallas Public Schools Education Committee, “TEA Position on Using TAAS Data and TAAS-Norm-Referenced Comparisons,” Dallas, TX, February 9, 1999: 1.

virtually every other state. On the 1998 eighth grade NAEP writing exam, the percentage of Texas Latino students scoring at the NAEP's "proficient" level was twice the national average, while the percentage of African American students scoring at the "proficient" level was nearly three times the national average.³⁸ A Business Roundtable-Education Trust report calculated that if African American fourth graders in every state scored as well in math as those in Texas, the national achievement gap between white and African American fourth graders in math would close by a third. The report also calculated that if African American eighth graders everywhere wrote as well as their peers in Texas, the national achievement gap between white and African-American eighth graders in writing would be cut in half.³⁹

Two well-publicized Rand Corporation reports have offered seemingly contradictory conclusions on the TAAS and the NAEP. A June 2000 Rand report found that states like Texas that had extensive accountability systems had the highest and most improved NAEP scores. That report said that the TAAS was an important factor in improving NAEP-measured academic achievement.⁴⁰ An October 2000 Rand report, however, questioned the validity of TAAS scores by showing that those scores did not correlate with the results of other standardized tests. Also, in the 1990s, Texas students' TAAS scores increased more than their NAEP scores.⁴¹

Harvard researcher Jay P. Greene points out that "It is possible that TAAS, which is based on the mandated Texas curriculum, tests different skills than those tested by the

³⁸Craig D. Jerald, *op. cit.*, 17.

³⁹*Ibid.*, 18.

⁴⁰See David W. Grissmer, Ann Flanagan, Jennifer Kawata, and Stephanie Williamson, "Improving Student Achievement: What NAEP State Test Scores Tell Us," The Rand Corporation, June 25, 2000: 55, 99–100.

⁴¹See Stephen P. Klein, Laura S. Hamilton, Daniel F. McCaffrey, and Brian M. Stecher, "What Do Test Scores in Texas Tell Us?" Rand Education Issue Paper, October 2000.

national standardized tests. Both could produce valid results and be weakly correlated to each other if they are testing different things.”⁴² Hoover Institution education economist Eric Hanushek notes that both Rand studies were based on poor research designs and that “neither holds up to a modicum of scrutiny.”⁴³ Greene, who contends that Texas has in reality made remarkable education gains, nonetheless observes that there is “some ambiguity” regarding any conclusions that can be drawn from a comparison of NAEP and TAAS results.⁴⁴

The controversy over the relationship between TAAS and NAEP scores aside, serious questions concerning the role and effect of TAAS in the Texas system still remain. Given the fact that TAAS assesses achievement at low levels, does TAAS put optimal pressure on schools to substantially improve achievement and learning? An analysis of the TAAS math exams by a team of mathematicians and statisticians associated with the math advocacy group Mathematically Correct concluded that “Low-level objectives are unlikely to bring student achievement up to the level of our international competition.”⁴⁵ Further, the study says

If the TAAS assessment levels are lower than optimal and there is a focus on minimum achievement relative to those assessments, there is an inherent risk that curriculum and instruction will be swayed toward sub-optimal levels as a result of the assessment process. Thus, it is possible that the TAAS exam system is not nearly as effective as it might be in promoting greater mathematics achievement statewide in Texas.⁴⁶

⁴²Jay P. Greene, “An Evaluation of the Florida A-Plus Accountability and School Choice Program,” Florida State University/Center for Civic Innovation at the Manhattan Institute/Program on Education Policy and Governance, Harvard University, February 2001: 3 (quotation edited to remove an unnecessary comma).

⁴³Eric A Hanushek, “Rand versus Rand: The Sequel,” *Education Matters*, Spring 2001: 69.

⁴⁴Jay P. Greene, *op. cit.*, 3.

⁴⁵Paul Clopton, Wayne Bishop, and David Klein, “Statewide Mathematics Assessment in Texas,” <<http://www.mathematicallycorrect.com/lonestar.htm>>, 29.

⁴⁶*Ibid.*, 20.

The analysis concludes that the emphasis on minimum achievement on TAAS results in a flawed accountability system:

[T]he system of mathematics achievement assessment in Texas emerges as a powerful model but one that is too highly focused on minimal achievement. The incentives for improvement that accompany the statewide assessment system do not emphasize high achievement sufficiently. In fact, the design of the assessment devices themselves doesn't even permit the measurement of high achievement levels with any degree of accuracy. Without a substantial adjustment to the objectives that are evidenced by the exam items themselves, it seems unlikely that the assessment system will effectively promote the kind of achievement necessary for students to realize the full benefit of a rigorous mathematics education.⁴⁷

A 2000 study prepared by Laurence A. Toenjes and Jean E. Garst for the state department of education supports the conclusion of the Mathematically Correct group. Though not indicting the entire TAAS-based accountability system, the Toenjes-Garst study notes that “a great many students who do well on the TAAS later do poorly on the Algebra I End-of-Course test.”⁴⁸ Thus, for example, “of eighth-graders who earned a Texas Learning Index (TLI) score of 80 on the math portion of the TAAS only about 27 percent passed the end-of-course algebra test later as ninth graders.”⁴⁹

⁴⁷Ibid., 31.

⁴⁸Laurence A. Toenjes and Jean E. Garst, “Identifying High Performing Texas Schools and School Districts and their Methods of Success,” prepared for the Texas Education Agency, December 2000: i.

⁴⁹Ibid. Clopton points out: “Although the Algebra 1 [EOC] exam appeared to be at a notably low difficulty level, the jump in grade level achievement required between the 8th- or 10th-grade exams and the Algebra exam was striking, roughly two grade levels by California standards. This suggests that the preparation for algebra, as measured by TAAS exams, is likely to be insufficient, leaving students who pass their TAAS exams at a risk of failure in algebra. . . . Performance criteria in the grades leading up to algebra are simply not sufficient to support success.” See Clopton, *op. cit.*, 57–58.

Thus, although increased learning is taking place in Texas, the incentives and pressures created by TAAS may not be promoting as large an improvement as possible.

The TAAS exams have also suffered from changing difficulty levels from year to year. Harvard researcher and standards expert Sandra Stotsky found that “The 1995 [reading TAAS] tests are longer and more difficult than the 1998 tests at all grade levels.”⁵⁰ Stotsky concludes that “If the scores students achieved on the 1998 tests were higher than those achieved by their counterparts on the 1995 tests, the decline in the overall level of reading difficulty of the selections on these tests . . . suggests that there may have been no real improvement in their reading skills.”⁵¹ Indeed, Stotsky says that “There may have even been a decline.”⁵²

Because the TAAS exams are criterion-referenced, they are supposed to be aligned to the state’s standards and curriculum. The Texas academic content standards have received generally good reviews from evaluators. The Thomas B. Fordham Foundation gives the state’s standards an overall grade of B, with specific grades of B for the English and math standards.⁵³ The current standards, the Texas Essential Knowledge and Skills (TEKS), replaced the old Texas Essential Elements in 1998–99. The new standards, evaluators say, are more rigorous than the old, so a test based on new ones should be more challenging than an exam based on the old ones. The state is creating a new TAAS exam, dubbed TAKS, aligned to the new standards, and it will be used in 2003. (In 2004, a new TAAS eleventh grade exit exam will replace the current tenth grade exam. It will cover not only language arts and math but also several other core subjects.)

Skeptics point out, however, that as the test becomes more rigorous, Texas’ performance criteria will require lower

⁵⁰Sandra Stotsky, “Analysis of the Texas Reading Tests, Grades 4, 8, and 10, 1995–1998,” unpublished paper, 1998: 8.

⁵¹Ibid.

⁵²Ibid.

⁵³“The State of Standards 2000,” *op. cit.*, 115.

cut-scores for passing, which will mean that “the utility of this increase in rigor for stimulating greater achievement is questionable.”⁵⁴ Further, an independent review of the proposed TAKS, done in cooperation with the state department of education, found that the new TAKS objectives were poorly delineated, were repetitious from grade to grade (with no increase in difficulty from grade to grade), did not reflect the expectations set forth in the new state standards, lacked academic rigor, and covered too few topics too superficially.⁵⁵

For whom are schools accountable in Texas? Originally, only regular education students had to take the TAAS exams. In 1999, the state administered the exam to learning-disabled students and non-English-speaking students (the latter taking the TAAS in a Spanish-language version in various grades). Still, many students receive exemptions from taking the test. In 2000, 13 percent of African American students, 12 percent of Hispanic students, and 14 percent of economically disadvantaged students were not tested, versus 7 percent of whites.⁵⁶ Either schools labeled these students learning-disabled or non-English-speaking, or the students were absent, or their school had some other reason for not testing these students. It should be noted that, since 1991, enrollment of the learning-disabled in Texas has increased by 32 percent, with much of that increase consisting of minority students. As Judson points out, those who have fallen outside the

⁵⁴Paul Clopton, *op. cit.*, 59. Clopton cites the example of the Algebra 1 TEKS: “The limitations of the TEKS with respect to quadratics are consistent with those throughout the Algebra 1 TEKS. Algebraic methods are de-emphasized, especially when difficult. They are also unnecessarily vague, making students, teachers, and test-designers unsure as to the exact expectations. These characteristics may make higher test scores possible, but they will not promote greater achievement in algebra.” *Ibid.*, 61. Clopton concludes that “the shift to a TEKS-based TAAS can provide only a slight improvement at best.” *Ibid.*, 59.

⁵⁵See Chris Patterson, testimony to Texas Board of Education, Texas Public Policy Foundation, July 12, 2001.

⁵⁶Craig D. Jerald, *op.cit.*, 10.

accountability system are “ironically the students who most need accountability from the system.”⁵⁷

The ease with which schools can meet the performance standard needed to achieve an Acceptable rating in the Texas system has perverse consequences. Because only 50 percent of students at a school have to pass the TAAS in order for the school to gain Acceptable status, schools often focus reduced attention on students who are likely to pass the TAAS and give additional attention only to students who are less likely to pass the exam.⁵⁸ Given the fact that the TAAS is not as difficult or rigorous as other tests, many students who can pass TAAS, but not other tests, may not receive as much attention as they should be getting.

Is progress measured absolutely or relatively? In Texas, accountability ratings are based on absolute measures of the percentage of students passing TAAS exams, dropout rates, and attendance rates. Progress is measured through the use of the student growth-based Comparable Improvement indicator.

Is Texas’ accountability system fair? Texas law requires that TAAS scores be reported separately for African American, Latino, white, and low-socioeconomic students. The Texas accountability system then requires that in order to achieve a given rating level (e.g., Exemplary, Recognized, etc.), not only must a school meet the test-score target, but all subgroups of students must also meet that target. Thus, for example, to achieve an Acceptable rating, at least 50 percent of all students at a school must pass the TAAS and at least 50 percent of each subgroup of students at the school must also pass the TAAS. No group of students, therefore, can be ignored under the Texas system.

The gap between the passing rates for white and minority TAAS test-takers has decreased significantly from 1994 to 2000. For instance, on the tenth grade exit exam, the gap between whites and African Americans has closed from 36

⁵⁷Jeff Judson, *op. cit.*, 34.

⁵⁸*Ibid.*, 33.

percent in 1994 to 22 percent in 2000. On the same exam, the gap between whites and Latinos has closed from 30 percent in 1994 to 19 percent in 2000.

Caution should be used in interpreting these improvements in minority passing rates. If the TAAS is not as difficult as it should be and if state officials have reduced the performance level needed in order to pass the TAAS, then one could expect less room for whites to improve on the test, since they started at a relatively high achievement level. At the same time, minorities, since they started at a lower achievement level, would have more room to improve. When whites improve modestly and minorities improve greatly, this result may seem impressive. But the narrowing of the gap in achievement scores may (or may not) have more to do with the difficulty of the test and how it is scored than with a narrowing of the gap in actual achievement.

Is the Texas accountability system informative? As in the case of California, the answer is yes and no. The state makes available (to parents and the public) school report cards containing TAAS passing rate information, attendance and dropout rates, plus an array of other information. However, the state does not make generally available the longitudinal data on student achievement that would make it possible to evaluate the effectiveness of individual teachers based on value-added calculations.

What about Texas' rewards-and-sanctions program? The various rewards and sanctions have increased pressure on schools and districts to improve their accountability ratings. A study (sponsored by the state department of education) surveyed school and district administrators and found a great deal of pressure to improve. One local superintendent said that his district allows new principals three years to move their school up to a Recognized rating level. For principals of schools that have been awarded a Recognized rating but then slide back below Recognized in a given year, this district gives the principal one year to return the school to its previous Recognized status. In another district, principals

are told that if their school is not at the Recognized level by the close of the school year, they should not anticipate a renewal of their contract.⁵⁹

Further, the study sponsored by the state department of education found that all middle schools and high schools surveyed had teams working to align the math curriculum with the new Texas standards. What schools were doing varied from school to school. Some schools held general discussions, whereas others made an in-depth effort to coordinate curriculum designs.⁶⁰

There are problems, though. When researchers for this study asked whether districts had a continuous TAAS-based cycle of teaching, testing, data analysis, and reteaching, they found enormous variability across districts.⁶¹

Also, programs designed to assist low-performing students have serious irregularities. First, the criteria for determining exactly which students are considered low-performing differed markedly from school to school. For example, schools varied in criteria they used to categorize students as low-performing in mathematics. The Toenjes study cites criteria from five separate campuses:

1. Only students who failed the TAAS math test
2. Students who failed TAAS together with students failing a math class
3. Students who scored low on a standardized pretest, such as the ITBS, together with those who also failed TAAS
4. Only students who fail a math class during the year
5. Only students who are recommended by teachers⁶²

⁵⁹Laurence A. Toenjes and Jean E. Garst, *op. cit.*, 28.

⁶⁰*Ibid.*, 30.

⁶¹*Ibid.*, 40.

⁶²*Ibid.*, 34.

Thus, a student considered low-performing at one school is not considered low-performing at another (and is therefore not eligible for special assistance). Further, the study found inconsistency in record-keeping from school to school. Some schools closely watch the scores of low-performing students, whereas at other schools there is no focus whatsoever on the ongoing performance of these students.⁶³

Local school administrators also said that rules and constraints imposed by electives and sports made it impossible to have mandatory programs for low-performing students in math. Although all the districts surveyed do provide programs that students can sign up for to get help in math, many of the programs are offered for only part of the year (usually the second semester in preparation for the spring TAAS).⁶⁴ Of course, this means that many students end up not getting assistance either because they decide not to take advantage of the voluntary program or because the program is not offered at a particular time of the year. Even where low-performing students voluntarily attend after-school or tutoring classes, the study found that “usually no single teacher or administrator has responsibility for monitoring their progress to see if any real gains are being made.”⁶⁵ Indeed, schools simply go through the motions in their remedial activity. The result is: “Whether or not the student actually is making progress is not discovered until the next TAAS testing. Therefore, many low-performers become chronic failures.”⁶⁶

Students do have an incentive to do well enough on the tenth grade TAAS exit exam to earn their high school diploma. However, rather little happens to students who don't do well on other TAAS tests.

As for teachers, the state accountability system does not formally tie rewards or sanctions to individual teacher

⁶³Ibid.

⁶⁴Ibid.

⁶⁵Ibid.

⁶⁶Ibid.

performance. But in a right-to-work state with union rules that are less encumbering than those in other states, local school officials may have more freedom to address individual teacher performance. (For this, it would be helpful if the state made value-added data generally available.) Also, as previously noted, many districts are willing to fire principals who do not improve their school's accountability rating.

Although the state allows students attending a poor-performing school to transfer to a better-performing public school, the state does not tell districts to fund exit vouchers for such students to attend private schools, nor does the state fund such vouchers itself. In the absence of exit vouchers, students from poor households are likely to remain in the public school system in comparatively weak schools. Because, in most cases, students cannot use an exit voucher to leave the public system and take funding with them, public school officials have less incentive to improve student achievement.

In sum, there is certainly good news in Texas. Even critics acknowledge that because the state set up its accountability system, students are learning more than they did before. Nonetheless, many problems remain, and the state needs to address them, especially in the area of testing. Texas officials have said from the outset that their system was an evolving one. Time will tell if the final product is worth the growing pains.

FLORIDA

In 1999, Florida enacted the A-Plus accountability program. The program includes a number of unique and innovative features and, as a result, has been widely discussed. Prominent features of the system include a new standards-aligned test, school grading, merit pay for teachers, and exit vouchers for students in failing public schools.

Results from the Florida Comprehensive Assessment Test (FCAT) are the cornerstone of the A-Plus accountability

program. The test contains two basic components: a portion measuring selected benchmarks in reading, writing, and mathematics based on the Sunshine State Standards, the state's academic content standards; and, from 2000–01, a second part measuring each student's performance against national norms. Thus, the test contains both criterion-referenced and norm-referenced sections.

Beginning in 2000–2001, FCAT tests grades 3–10 (previously, only grades 4, 8, and 10 were tested in reading and grades 5, 8, and 10 in math). The FCAT reading exam assesses students' ability to construct meaning from informational text and from literature. The FCAT math exam assesses students in six areas, including number sense, concepts and operations, measurement, geometry and spatial sense, algebraic thinking, and data analysis and probability. FCAT includes multiple-choice questions and constructed-response questions (questions where students write an answer, solve a problem, give an explanation, or draw a sketch).

FCAT has five grades (also called performance-standards levels), Levels 1 to 5 (with 5 being the top level). The definitions of the levels are:

- Level 5: The student has success with the most challenging state content standards and answers most of the test questions correctly, including the most challenging ones.
- Level 4: The student has success with the challenging state content standards and answers most of the test questions correctly but may have only some success on the most challenging ones.
- Level 3: The student has partial success with the state standards but performs inconsistently, answering many of the questions correctly but having less success on the most challenging ones.
- Level 2: The student has limited success with the challenging content of the state standards.

- Level 1: The student has little success with the state standards.⁶⁷

Student FCAT scores range between 500 for Level 5 and 100 for Level 1.⁶⁸

In addition to the FCAT, Florida has two other state tests: a writing exam and a high school exit exam. The writing exam, Florida Writes!, is given in grades 4, 8, and 10. The exam is performance-based and criterion-referenced. It has achievement levels ranging from 6.0 (high) to 1.0 (low). In California, students who do poorly on the state-sponsored standardized tests face little in the way of consequences, but in Florida, students must meet standards (as assessed by FCAT and the writing exam) in order to be promoted to the next grade.

The High School Competency Assessment Test (HSCAT) is a criterion-referenced exit exam administered in the eleventh grade. A student must pass the reading, writing, and math sections of HSCAT to receive a high school diploma. Florida is phasing out the HSCAT. Students can currently substitute the tenth grade FCAT for the HSCAT, and by 2003, the HSCAT will be discontinued in favor of the FCAT.

Under the A-Plus plan, the state grades schools on an A–F scale. In December 2001, Gov. Jeb Bush and the Florida cabinet approved changes to the original criteria for the various grades. The new grading system, which state officials believe will slightly increase the number of schools receiving As and Fs, will use a point system based on student performance on the FCAT. Under the new grading formula, a school's performance will be measured in three ways:

- The percentage of students scoring at the highest three levels on the reading, math, and writing tests.

⁶⁷Margaret Goertz and Mark Duffy, *op. cit.*, FL 3.

⁶⁸Holly Stepp, "Gov. Bush, Cabinet OK New School-Grading Rule," *Miami Herald*, December 19, 2001, 1B.

- The individual progress made by each student over his or her previous year's FCAT score.
- The improvements in reading scores made by the lowest-performing 25 percent of students.⁶⁹

Specifically, schools will earn points that will determine their grade based upon various student performance indicators:

- One point for each percent of students who score at Level 3 or higher in reading. So, if 25 percent of students at a school score at Level 3 or higher in reading, the school receives twenty-five points.
- One point for each percent of students who score at Level 3 or higher in math.
- One point for each percent of students who score at Level 3 or higher in writing, then averaged with the percent of those who score at Level 3.5 or higher.
- One point for each percent of students who make annual gains in reading.
- One point for each percent of students who make annual gains in math.
- One point for each percent among the lowest 25 percent of students at a school who make annual gains in reading.⁷⁰

After adding up all the points, a school's final grade will be based on the following scale:

- A: At least 410 points
- B: At least 380 points
- C: At least 320 points
- D: At least 280 points
- F: Less than 280 points⁷¹

⁶⁹Ibid.

⁷⁰Ibid.

⁷¹Ibid.

Under the previous grading system, it was possible for schools to count on their higher performing students to mask the low performance of other students. Florida officials believe that the new grading system, which emphasizes individual student performance, will force schools to address the achievement problems of low-performing students. Gerry Richardson, director of evaluation and reporting for the Florida Department of Education, points out: "If before schools were counting on their highest-achieving students to make up for the ones who were struggling, that is no longer acceptable. Schools will have to make sure that all of their students are making progress."⁷² Adds Betty Coxe, deputy commissioner of educational programs: "There shouldn't be a principal or teacher in the state saying, 'Oh, now that the state is going to grade us on whether all our students can read, we'll actually teach them all to read.' It should have been a priority all along."⁷³ The new grading system's increased focus on individual student performance, therefore, will act as an incentive for schools to improve the achievement of all students.

Florida has not set "passing" scores for any grade level for the FCAT except for grade 10. The state simply reports scores on a scale from 100 to 500 and in terms of the achievement levels (with cut-scores for each level). The state provides FCAT results for each individual student, school, and district, as well as statewide data. The state reports a total score for each student on each test. The state also reports the student's performance on different strands of academic content and in terms of the five achievement levels. In addition, the state reports the student's ranking within Florida itself and as compared to national norms. For schools, districts, and the state overall, the state produces A-Plus reports showing average scores and the percentage of students performing at the five achievement levels.⁷⁴

⁷²Ibid.

⁷³Ibid.

⁷⁴Florida Department of Education, "FCAT Briefing Book," Tallahassee, FL, February 2001: 4, 8.

Florida plans to use the FCAT results for value-added analysis, which will report an individual student's progress from one year to the next. Value-added analysis, which Tennessee uses in its accountability system, provides statistical measures of the influence that school districts, schools, and teachers have on student learning. According to Tennessee education officials, one of the most powerful and controversial aspects of value-added assessment "is that it can reach beyond the school level to produce a measure of an individual teacher's effectiveness, based on how well the students in his or her classroom perform each year."⁷⁵ The Florida Department of Education seems to agree, noting, "The progress of all students in a school can be reported in terms of individual teachers who provide instruction to those students."⁷⁶ As a consequence, individual teachers can be held accountable for how much they add to the learning of their students.

Florida issues a School Accountability Report that grades each school using the A–F scale. The report also includes the percentage of students scoring at each achievement level within a school, plus a school's suspension rate, absentee rate over twenty days, dropout rate, promotion rate, percent receiving free-or-reduced lunch, and mobility rate.

The A-Plus program provides a number of rewards for schools and teachers. Schools qualify for additional funding if they meet the higher performing grade A criteria, show significant improvement, or improve by one letter grade from one year to the next. The funding is allocated on a per-student basis.

In 1997, the Florida legislature required school boards to base a portion of each employee's compensation on performance. Hence, local superintendents must propose a schedule for teachers' salaries that is based in part on student performance. Local districts decide on these

⁷⁵Margaret Goertz and Mark Duffy, *op. cit.*, TN 13.

⁷⁶"FCAT Briefing Book," *op. cit.*, 5.

salary schedules, and the state does not mandate how much of a teacher's salary must be based on the achievement of students. Revisions passed in 1999 require that evaluations of teachers and administrators be based on the performance of students assigned to their classrooms or schools. In other words, local administrators are to judge teachers based on their individual performance in the classroom, not as part of a collective schoolwide body of teachers.

Also in 1999, when the legislature passed A-Plus, it mandated that teachers and administrators who demonstrate outstanding performance (based upon student achievement) could earn annual bonuses of up to 5 percent of their base pay. The bonuses under this unique merit-pay system will be based on FCAT results and will begin in 2002. Some teachers have worried that this legislation could be interpreted as also implying a 5 percent pay cut for teachers at low-performing schools. Governor Jeb Bush, however, in a teleconference with Florida teachers, has repudiated the pay-cut interpretation.⁷⁷ Florida's planned value-added FCAT measurement system will be a useful tool in rewarding individual teacher performance through these merit bonuses.

Low-performing schools are subject to a variety of sanctions. A school or district receiving a D or F grade is eligible for the following: (1) state intervention, assistance, and funding; (2) a community assessment team to make recommendations for intervention and assistance to improve the school's performance; (3) priority to receive technical assistance and training services from the state; (4) priority in the use of state supplemental funds; (5) district intervention and assistance; and (6) district authority to declare an emergency and to negotiate special provisions to free the schools from contract restrictions that limit the school's ability to improve student performance.

⁷⁷"Bush Defends A-Plus Plan, Says Teacher Bonuses Misunderstood," *Florida Sun-Sentinel*, April 10, 2000.

If a school has received an F grade in any two years of a four-year period, students at that school may: (1) attend a higher performing public school in the district, (2) attend a higher performing public school in an adjacent district as long as there is space available, (3) stay at the same public school, or (4) receive an exit voucher to attend a private (secular or nonsectarian) school. We will discuss this latter option, the so-called “voucher sanction,” in greater detail below.

For what are Florida schools accountable? The FCAT is a criterion-referenced exam based on the state academic content standards. The Thomas B. Fordham Foundation has given Florida’s English-language arts standards a B grade and given a disappointing grade of D to the state’s math standards. The Fordham reviewers criticized the math standards for having expectations that “are too often very low indeed.”⁷⁸ A statewide test based on such standards might, then, also have expectations that are too low.

On the other hand, Greene, in his study of the A-Plus program, found that, unlike TAAS, FCAT stacked up reasonably well against standardized tests like the Stanford-9, which Florida students took under low-stakes circumstances in spring 2000. According to Greene:

In the second Rand Corporation study of TAAS in Texas, Stephen Klein and his colleagues never found a correlation of more than .21 between the school level results from TAAS and the school level results of low stakes standardized tests. In this analysis, we never found a correlation between FCAT and standardized tests below .86. All of these correlations in Florida are statistically significant, meaning that the strong relationship between the results of the two tests is very unlikely to have been produced by chance.⁷⁹

Greene concluded that such numbers support the validity of the FCAT reading and math scores and that “Schools in

⁷⁸“The State of Standards,” *op. cit.*, 44.

⁷⁹Jay P. Greene, *op. cit.*, 5–6.

Florida perform on the high stakes FCAT similarly to how they perform on the low stakes Stanford-9.”⁸⁰

Greene’s findings have to be understood in context. His study is only trying to see if Florida’s schools responded to incentives. He wanted to know if, faced with rewards and sanctions, Florida’s scores went up or not. One should not assume from his study that Florida’s standards are world-class, that the FCAT or the Stanford-9 is challenging, or that the state of Florida’s cut-scores represent performance at the levels of top-performing countries.

For whom are schools in Florida accountable? Virtually all students must take the state tests. The state excludes non-English-speaking students with fewer than two years of English-as-a-second-language instruction but tests those with more than two years of such instruction. On the other hand, the state allows non-English-speaking students who are tested to have extra time to divide the test into shorter testing periods or to have the teacher read questions out loud, or combinations of these accommodations. Some learning-disabled students may also be exempted, and those learning-disabled students who do take the test are given a modified form.

Is progress measured absolutely or relatively? Thus far, Florida has measured student progress in terms of reaching target scores that correspond to achievement levels. However, when Florida’s value-added measurement system is in place, the state will be able to track relative gains made by teachers, schools, and districts.

Is Florida’s accountability system fair? The state breaks down performance data into subgroups (economically disadvantaged, African American, white, Hispanic, Asian, and American Indian). In A or B schools, no subgroup can perform below the minimum criteria. However, Florida does not require that subgroups perform at a certain level in order for a school to receive a C grade (unlike Texas, which

⁸⁰Ibid., 6.

requires that a certain percentage of each subgroup perform at a set level for the school to obtain an Acceptable rating).

Furthermore, the exit voucher option for students at failing schools allows some students from poor households to escape from a failing public school to a better-performing private school. Parents of most of these students could not usually afford to send them to private schools.

Is Florida's system informative? The state distributes individual student scores and the data on school report cards to districts, schools, and parents. State and district data are available on the Internet. It will be interesting to see, once value-added measurement is adopted, whether Florida will distribute information on individual teacher performance to parents. In Tennessee, where value-added analysis is part of that state's accountability system, the state provides value-added information on individual teacher performance only to school officials and to the teachers themselves.

What about rewards and sanctions in Florida? On the reward side, the state's merit-pay system is the most promising feature. Teachers' unions have fought merit pay across the country. If Florida's merit-pay system proves successful, it could be an example that inspires similar programs in other states and districts. Perhaps even more important, if value-added analysis is folded into the system, then Florida's merit-pay program would be based on an objective indicator of individual teacher performance and would give individual teachers incentive to improve their teaching and focus on teaching methods that actually increase student achievement.

On the sanctions side, the exit voucher option has fostered the most controversy but offers the most promise. Although few students have become eligible for the state-funded exit vouchers, there appears to be a positive "voucher effect" on the performance of failing public schools. In his analysis of A-Plus's exit voucher program, Greene found that "schools that received F grades in 1999 experienced increases in test scores that were more than

twice as large as those experienced by schools with higher state-assigned grades.”⁸¹ The implication, of course, is that schools threatened with vouchers worked harder to improve their performance.

In response to the counterargument that the gains at failing schools had causes other than the prospect of vouchers, Greene compared higher-scoring F schools and lower-scoring D schools. He found that the F schools made higher gains than the D schools. Greene observed that “Given that the higher-scoring F schools were very much like the lower-scoring D schools, the fact that those schools that faced the prospect of vouchers made larger gains suggest that vouchers provide especially strong incentives to public schools to improve.”⁸²

Greene says that two forces are working to motivate schools to improve. First, all schools want to avoid the embarrassment of poor FCAT scores. But, said Greene, “schools with F scores had a second and very strong incentive to improve to avoid vouchers.”⁸³ That incentive, according to Greene, was the prospect of market competition:

Companies typically anticipate competitive threats and attempt to make appropriate responses to retain their customers before competition fully materializes. Similarly, it appears as if Florida schools that foresee the imminent challenge of having to compete for their students take the necessary steps to retain their students and stave off that competition.⁸⁴

Concrete steps taken by local school officials have included implementing traditional teaching methods (such as direct instruction and drill and practice) in Lake County, switching to a phonics-reading program in Miami-Dade County, and requiring Saturday tutoring in Broward County. A report co-sponsored by the Miami Urban League

⁸¹Ibid.

⁸²Ibid., 8.

⁸³Ibid., 11.

⁸⁴Ibid., 8–9.

and several think tanks observed that “the important thing is that the [exit voucher program] has instilled in the public schools a sense of urgency and zeal for reform not seen in the past when a school’s failure was rewarded only with more money that reinforced failure.”⁸⁵

Greene, however, cautions that the A-Plus program is still new and that it may change for better or worse in the future. So far, students at only two schools have received exit vouchers. If the number of students eligible for the vouchers does not grow in the future, Greene warns that “it is possible that the prospect of having vouchers offered to students will not seem so imminent to schools and they will not face the same incentives to improve.”⁸⁶ Perhaps Florida’s criteria for offering exit vouchers should be more inclusive to allow more students the possibility of receiving them, thus keeping up the pressure on public schools to improve.⁸⁷

Compared to California and Texas, Florida’s accountability system includes programs that offer better incentives for schools and school personnel to reform their ineffective ways and to improve student achievement. It will be up to Florida policymakers to follow up on the state’s promising beginning and to not lose sight of the principles and goals that put the system on the cutting edge of the accountability movement.

CONCLUSION

Despite the holes in many accountability systems, it is important not to lose perspective. Even under a less-than-optimal

⁸⁵As cited in Lance T. Izumi, “School Choice Improves Public Schools,” Pacific Research Institute for Public Policy, *Capital Ideas*, vol. 5, no. 43, October 25, 2000.

⁸⁶Jay P. Greene, *op. cit.*, 9.

⁸⁷The December 2001 changes in the state’s system of grading schools were brand-new at the time this was written, and their precise effects were unknown.

accountability system, the situation is often much better than when no accountability system existed. Not long ago, for example, California had no statewide assessment, had no sanctions for poor-performing schools, and paid little attention to improving student achievement. California’s accountability system has focused the attention of adults in the public school system—from teachers to principals to superintendents to local and state policymakers—on the priority of raising student performance. Because of that attention, better curricula have been adopted, better teaching methods are entering the classroom, and pragmatism about what works is replacing faddish ideologies.

Nonetheless, there is much room for improvement. In California, too many low-performing schools are not subject to sanctions. In Texas, a state test that has not been challenging enough may have allowed too many schools to escape sanctions. In Florida, the voucher sanction may not apply to enough schools. State officials and accountability advocates must address these and many other problems if accountability systems are to reach their full potential.

The principles of a democratic society include responsibility on the part of public servants. In a democratic society, public institutions, including public schools, must be accountable for their results. Only when citizens can find out how their tax dollars are being used are they in a good position to demand change. Although serious accountability in public schools is only in its infancy, the movements in this direction across the country are encouraging. After decades of promises, we will now have incentives and accountability that can bring real improvement in student achievement.

ABBREVIATIONS

- FCAT.....Florida Comprehensive Assessment Test
- HSCAT.....High School Competency Assessment Test
(State of Florida)
- HSEE.....High School Exit Examination (State of California)

- ITBS.....Iowa Test of Basic Skills (Riverside)
- NAEP.....National Assessment of Educational Progress
- Stanford-9.....Stanford Achievement Test, 9th ed. (Harcourt)
- TAAS.....Texas Assessment of Academic Skills
- TAKS.....Texas Assessment of Knowledge and Skills