



Texas AFT

A Union of Professionals

www.texasaft.org

**Beyond TAKS (and NCLB):
Putting Texas School Accountability
Back on Track**

Discussion Draft, October 3, 2008

TABLE OF CONTENTS

Executive Summary

Introduction

Systemic Flaws

The Proper Place of Testing

Guidelines for Reform

- Better Standards as a Foundation
- Measuring Student Gains—Accurately
- Acknowledging the Limits of Any One Test
- Supportive Interventions
- Appropriate Testing of Students With Disabilities
- Appropriate Testing of English Language Learners

An AFT Alternative to the Current System

- Academic Achievement
- Safe Learning Environment
- Teamwork to Promote Achievement
- Resources
- A Spur to School Improvement
- A Balance of Quantitative and Qualitative Accountability Criteria

Implications for Use of Test Scores in Professional Evaluation

- A Look at Current Experiments
- Lessons From Business Practices

Control Over Instructional Time

Factors Outside the Schoolhouse

Independent Review of Any New System

Conclusion: A Vision of Accountability Based on Classroom Realities

Executive Summary

A grass-roots revolt is brewing in reaction to the ill effects of test-driven school accountability under the Texas law that inspired the No Child Left Behind Act. Texas appears to be a pioneer once again, this time in contemplating a major course correction for its accountability system. After more than a dozen years of experience with the growing abuse of testing under that system, it is time to draw necessary lessons and completely rethink current policy.

The current system of test-based accountability is deeply flawed. Despite claims that testing has driven school improvement, the reality is that improvement has been too limited and cannot be attributed to testing in isolation from other factors.

Counterproductive practices associated with the current system include:

- narrow, basic-skill or even below-basic tests that fail to assess higher-level knowledge and skills;
- narrowing the curriculum to fit the test;
- unfocused, weak, incoherent state standards that fail to define what is most important for students to know and be able to do at each grade level and in each subject;
- excessive time devoted to testing and test preparation, so that teachers lack time to provide real instruction;
- contradictory school ratings that tell us little or nothing about school effectiveness;
- destructive emphasis on punitive sanctions instead of supportive interventions for struggling schools;
- overidentification of schools as failing, so that corrective action is not focused where most needed;
- failing to challenge students at every level of achievement to perform at higher levels;
- failure to give credit for students’ academic gains, only for passing a fixed “proficiency” hurdle;
- misuse at the same time of unproven “growth” measures for high-stakes decisions;
- misplaced reliance on a single, imprecise test when multiple measures are needed;
- pressure to test students with disabilities in ways that are developmentally inappropriate;
- pressure to test English Language Learners inappropriately in English, thereby inaccurately assessing their subject-matter knowledge;
- utter neglect of the role of factors outside the schoolhouse in determining school performance.

A sound alternative to the current system should take into account not just state test scores but all factors known to affect student achievement. An alternative model proposed by the United Federation of Teachers, AFT’s local affiliate in New York City, would replace one-dimensional accountability based on test scores with a multi-factor approach that separately grades each school in four dimensions: academic achievement (with state assessments serving only as one of

several measures of achievement); safe learning environment; teamwork to promote achievement; and resources provided by state and district to meet students' needs.

The point of this alternative approach is not to fix blame but to fix schools. This approach demands extensive, detailed, and independent qualitative as well as quantitative evaluation of school and system-wide effectiveness. It entails a big investment in independent expertise as well as in improved standards and tests based on those standards that measure higher-order achievement.

Devotees of the current, test-based accountability system with its misleading appearance of precise measurement will say that this multi-factor alternative approach aims to take the pressure off of schools, to excuse low performance. The opposite is true. This alternative "balanced scorecard" approach, widely used in the business world, would expose all the players in public education—from the schoolhouse to Austin to Washington—to increased pressure to deliver on the promise of school improvement.

In place of the current, top-down, bureaucratic model of accountability, the alternative pictured here would make policy-makers and schools alike more accountable to parents and taxpayers at the grass roots who want schools and classrooms that serve every student well. This alternative also would restore the central role of teachers exercising professional judgment in providing all their students with instruction tailored to their individual needs.

Introduction

The modern standards-based educational reform movement began in the 1980s with the support of many teachers—particularly with the support of the American Federation of Teachers under the leadership of Al Shanker. The vision: a demanding curriculum built on strong, clear, sequential standards, enabling teachers to prepare their students for state assessments without "drill and kill."

Standards-based reform as originally envisioned meant that teachers and parents would get a better sense of what students are expected to know and be able to do at any specific grade level. It would help teachers identify which students are having difficulty and need extra help. It would allow teachers to develop, share, and refine best practices with their colleagues, and benefit from professional development based on the standards, not on the latest fads. Standards-based educational reform thus went hand in hand with the recognition of teachers as professionals engaged in challenging work, who should be at the center of a web of resources they could draw upon to enable every child to succeed.

Some real progress resulted. Texas helped to pioneer the disaggregated reporting of student performance, so that patterns of neglect of the educational needs of disadvantaged and minority

students would be exposed and corrected. The state also honored a tacit bargain with its educators—more demands for achievement from high-need students were accompanied, not always but often and substantially, by more resources for our schools to get the job done. Tangible examples cited by one prominent study of Texas achievement gains in the 1990s included limits on class size in grades K-4 and funding for kindergarten and pre-kindergarten. (Grissmer et al., 2000) Higher teacher pay and higher formula aid for school districts, more equitably shared, were implemented at the same time. State assessments were understood to be part of a bigger picture, not the be-all, end-all of education reform.

State assessments have since evolved, however, into the central focus of public education at both the state and federal level. Now concerns are rising about the ill effects of test-driven accountability standards under the Texas law that inspired NCLB and under that federal law itself. Proposals are proliferating for major changes in both state and federal accountability requirements for public schools.

Texas appears to be a pioneer again, this time in contemplating a major course correction for its accountability system. A consensus has emerged that it is time for strategic rethinking of the purposes, methods, and consequences of current accountability schemes—so much so that in 2007 the Texas House of Representatives voted to abolish the existing system entirely and replace it by 2009. That “sunset” provision eventually was amended to require an interim study and recommendations of needed changes, leaving the current system largely intact meanwhile. The Texas legislature’s Joint Select Committee on Public School Accountability accordingly is due to report recommendations by December 2008 that address, among other things, these central questions about the current system:

–“the extent to which the accountability system fairly and accurately reports the effectiveness of educators, instructional programs, support services, and financial expenditures and the impact of these elements on student achievement;”

–“the methods available to monitor the progress of each public school student, with special emphasis on methods to monitor demonstrable growth in academic achievement;”

–“the performance indicators that would successfully measure the effectiveness of the campus teaching and learning environment, including the effect of student discipline on that environment;” and

–“the extent to which the accountability system measures the performance of districts and campuses on important indicators and aspects of the educational process, other than student scores on standardized assessment instruments....”

Systemic Flaws

It is high time for this broad-gauged rethinking of the accountability system. Classroom teaching is increasingly distorted if not crippled by growing demands for test preparation, practice testing, benchmark testing, and test-data churning, leaving little time and few resources to apply diagnostic lessons from test results to help individual students learn. Testing mania has gone to absurd lengths. Teachers often are forced to become data-entry clerks, transferring data from one format to another for the convenience of central-office administrators. Teachers all too often are compelled to administer so many preliminary tests and practice tests tied to state assessments that they have no time to modify instruction based on the insights gleaned from the data gathered.

The standardized-testing system is defeating its own purpose, both dumbing down curriculum and impeding individualized instruction. As Chester Finn and Diane Ravitch, longtime advocates of testing for accountability, have written: “We did not foresee how completely standards-based reform would turn into a basic-skills frenzy.” (Rothstein, “Holding Accountability to Account,” National Center on Performance Incentives, February 2008) That frenzy has been compounded by sheer confusion, as federal and state accountability measures contradict each other, using the same test scores to label schools' performance simultaneously acceptable and unacceptable.

Advocates of the current test-driven accountability system contend that NCLB and TAKS-based accountability are responsible for improvement in student test scores. But their case is weak. There is evidence that achievement gains were greater before NCLB intensified the emphasis on TAKS exams than after, as measured by the National Assessment of Educational Progress. (citations forthcoming: NAEP, TAKS, TAAS data)

While a recent report on state test scores by the Center on Education Policy has been touted by federal officials as proof that those scores have risen because of NCLB, the study does not show that. Rather, the CEP analysis indicated only that increasing numbers of students are reaching state-specified proficiency levels, as defined in widely varying fashion from state to state, without independent corroboration of actual gains via NAEP results. The CEP report actually said that NCLB cannot be identified as the reason for test-score increases because researchers could not separate the impact of the 2002 law from the impact of many other initiatives that were being implemented at the same time. As Jack Jennings, president of the Center on Education Policy, said: “You have to be very careful....At the same time that NCLB was taking effect, a whole slew of things were happening....We cannot draw a direct line between this increase in achievement and NCLB.” (Education Week, “State Tests Show Gains Since NCLB; Report cautions against crediting education law,” June 6, 2007)

The same caution is in order regarding claims for the success of test-driven accountability in Texas, given the simultaneous implementation of many different education policy changes.

The appearance of significant gains on TAKS scores is not consistently corroborated by NAEP results. (citations forthcoming: NAEP, TAKS data) This should come as no surprise, because the definition of proficiency on TAKS tests falls at or below the “Basic” level on NAEP.

What is clear is that the system is flawed in many ways obscured by the seeming precision of the passing percentages and school and district ratings periodically reported in the newspapers. Just to name two very serious defects: The TAKS yardstick of “proficiency” at different grade levels is inconsistent; test difficulty varies too much from grade level to grade level to provide a useful gauge of students’ academic “growth.” Moreover, the TAKS-based testing system has been undermined by weak, repetitious standards from grade to grade for much of the core curriculum, so that passing the tests does not necessarily demonstrate progress toward true proficiency in the relevant knowledge and skills of each discipline.

Both federal and state lawmakers have compounded the problem by tying more and more high stakes to this fundamentally flawed test-based accountability system. The credibility of school accountability ratings has been seriously eroded, as federal and state ratings often contradict each other even though based on the same TAKS test results.

Both federal and state accountability ratings also discriminate against large, diverse districts, because they are subject to many more accountability trip-wires than smaller districts with fewer disadvantaged or special-needs students. Accountability indicators for student subgroups do not apply unless a district exceeds certain minimum numbers for students in such groups; smaller, more homogeneous districts are far less likely to meet those minimums and therefore far less likely to be held accountable for subgroup performance. A research brief by the Texas School Alliance notes that in 2006-2007 "all but one exemplary district were rated on fewer than 17 measures." (Texas School Alliance, “Confronting the Complexities of School Accountability,” 2007) In contrast, of the 33 larger school districts in the School Alliance, none were rated on fewer than 26 measures, and all 36 indicators were applied to large urban districts like Austin, Dallas, Fort Worth, and Houston. This inconsistent basis for school ratings has further eroded the accountability system’s credibility.

It is indeed time to start over, moving beyond both NCLB and TAKS to build a new system. It should be stipulated from the outset that the alternative accountability model outlined here will entail fundamental change not only in state law but also in aspects of the federal Elementary and Secondary Education Act, dubbed the No Child Left Behind Act in its most recent incarnation. Fortunately, the outlook for a major overhaul of that 2002 federal legislation, as early as next year, is encouraging, as more and more members of Congress recognize the limitations of test-driven accountability at the federal level. The balance of this report focuses on the state accountability system.

The Proper Place of Testing

The standards-based reform movement has gone badly off track. Ironically, in the name of improving achievement, we have ended up with too little focus on higher-level knowledge and thinking, too much focus on basic skills at a superficial level, not enough depth, not enough rigor. We force teachers to focus on superficial, basic-skills testing and test preparation, dwelling on narrow subsets of the wider curriculum our students need to master. For the sake of our students, teachers want to put the standards movement back on track.

It is well-established that the individual teacher is the most significant school-related factor in students' education. Many studies identify teachers as a crucial determinant of student achievement (Goldhaber and Anthony, 2004; Goldhaber, 2007; Rivkin, Hanushek, and Kain, 2005). As one scholar summed up the research consensus in testimony before Congress last year, "Quality teachers have a greater influence on pupil achievement than any other school-based factor." (Fallon, Carnegie Corp., May 17, 2007) A study of elementary schools in Austin, in St. Paul, Minnesota, and in New York City (Quint et al., 2007) sharpens the point, finding that "higher instructional quality is significantly related to higher student achievement in both reading and math. Instruction that requires students to tackle challenging tasks and to back up their assertions with evidence and reasoned arguments is especially associated with higher achievement." Yet in Texas we have systematically implemented an accountability system built around state testing that leaves less and less opportunity for teachers to exercise professional judgment in order to adapt instruction creatively to meet students' needs.

Make no mistake: Testing is an essential tool of teaching. Teachers use tests all the time to determine how much students know and can do, to diagnose individual deficiencies, to help individual students. Texas teachers are hungry for good data on their students' progress, and a key element of professional development for teachers is to learn how to use good data more effectively to help their students.

Teachers understand assessment, and they know how to evaluate students by formal and informal means. This is and should be the domain of the individual classroom teacher. The frenzy for standardized testing directly interferes with this important work, to the detriment of our students. (Lest anyone suppose teachers' own tests and grading decisions don't matter, please note: A student's GPA is one of the best predictors of college success.)

The issue, then, is not whether testing has a major role in teaching and learning. The issue is the misuse of state-mandated tests to make sweeping performance and productivity judgments about schools and even individual teachers. The fact is that the tests we have cannot serve this purpose because they reflect both out-of-school and in-school contributions to achievement and cannot distinguish between the two. A system that implicitly assigns all the responsibility for student achievement outcomes to individual schools and teachers is not a credible accountability system. And the harmful consequences are manifold.

Guidelines for Reform

The fundamental idea of using uniform state tests as one means to assess how much progress students and schools are making is not going to go away, nor should it. Parents and taxpayers deserve to know how their schools are doing. Disaggregating test results to show how well particular student subgroups are doing—by race, ethnicity, economic disadvantage, etc.—sustains pressure for equitable allocation of educational resources to reduce achievement gaps. These elements are worth keeping, but only as one part of a more meaningful report to stakeholders that truly accounts for the full range of meaningful variables contributing to student achievement.

Better Standards as a Foundation

To build a new system, Texas AFT starts from the premise that standards-based educational improvement must be just that: to make real educational gains we have to start with a solid foundation of meaningful, sequential standards. The point sounds obvious. Yet the whole structure of accountability under which Texas students and their teachers in the public schools have been operating has rested all along on a defective foundation of vague, superficial, incoherent standards in core subjects at many grade levels. Research by the American Federation of Teachers has found Texas standards defective at all levels for English language arts, in elementary school for science, and in elementary and middle school for social studies. (Glidden, “Common Ground: Clear, Specific Content Holds Teaching, Texts, and Tests Together, *American Educator*, Spring 2008, p. 19) The current process of standards revision undertaken by the State Board of Education thus far has been too haphazard and too politicized to remedy these defects.

Texas AFT believes we cannot get standards-based reform back on track unless we establish rich, sequential curriculum standards that are focused, rigorous, and coherent, so that tests aligned with the standards actually assess progress toward relevant knowledge and skills, not just test-taking prowess. Requiring additional courses in math and science for graduation from high school cannot make up for a lack of focus (superficial coverage of too many topics), lack of rigor (demanding too little in the way of higher-order, critical thinking skills), and lack of coherence (presenting topics with little concern for a progressive sequence that reflects the internal logic of each subject) in the teaching of core subjects. If we are serious about college readiness, students need to move much more deeply into fewer and more challenging topics in each subject in primary and middle school. (Schmidt, “What’s Missing From Math Standards? Focus, Rigor, and Coherence,” *American Educator*, Spring 2008)

As education scholar Linda Darling-Hammond has written, “most of our tests aren’t measuring the kinds of 21st century skills we need students to acquire and that are at the core of curriculum and assessment in high-achieving countries.” (Darling-Hammond, “High-quality standards, a

curriculum based on critical thinking can enlighten our students,” San Francisco Chronicle, October 14, 2007) Under the existing accountability schemes, Darling Hammond points out, “The plain truth is that the United States is falling far behind other nations on every measure of educational achievement. In the latest international assessments, the United States ranked 28th out of 40 countries in math—on par with Latvia—20th in science, and 19th in reading, even further behind than a few years ago. In addition, these other countries surpass us in graduation rates and, over the last decade, in higher education participation as well.” The rate of completion of post-secondary education is even lower.

Higher-achieving nations attain better results not because they all have national standards and tests—some rely on state-level standards and tests, some rely primarily on local assessment (and none use test scores as the basis for school accountability). They achieve these results, Darling-Hammond contends, because they have a curriculum focused on critical thinking, problem-solving, and examinations that require students to solve complex real-world problems and defend their ideas orally and in writing.

Darling-Hammond observes: “In most cases, their assessment systems combine centralized (state or national) assessments that use mostly open-ended and essay questions with local assessments given by teachers, which are factored into the final examination scores. These local assessments—which include research projects, science investigations, mathematical and computer models and other products—are mapped to the syllabus and the standards for the subject and are selected because they represent critical skills, topics and concepts. They are generally designed, administered and scored locally.”

The contrast with our mostly multiple-choice tests could scarcely be more marked. Where our state assessments typically test fact recognition over a wide array of topics within a subject, higher-achieving countries typically ask students to apply their knowledge and skills in a subject in ways that demonstrates their reasoning power.

Darling-Hammond cites a vivid case in point: “In the United States, a typical item on the 12th-grade National Assessment of Educational Progress, for example, asks students which two elements from a multiple choice list are found in the Earth's atmosphere. An item from the Victoria, Australia, high school biology test (which resembles those in Hong Kong and Singapore) describes how a particular virus works, asks students to design a drug to kill the virus and explain how the drug operates (complete with diagrams), and then to design and describe an experiment to test the drug—asking students to think and act like scientists.”

Another case in point right here in Texas, as recently reported in the Houston Chronicle (Rick Casey, “Commentary: Professor worries TAKS test breeds ignorant voters,” August 23, 2008): Dr. Kevin Jefferies of Alvin Community College has found that students entering college could score well on previous years’ TAKS questions without understanding basic concepts about American government or history. A majority of students could respond to cues built into TAKS

questions and come up with the right multiple-choice answer, yet when asked open-ended questions covering the same material were, literally, clueless.

The reality is that our testing regime is driving teachers to prepare students more for the skills they need to pass state tests than to prepare them for the real world in which they will be expected to apply what they have learned in school. Darling-Hammond cites this comment from one frustrated teacher: “I have seen more students who can pass the [state test] but cannot apply those skills to anything if it’s not in the test format. That observation mirrors the experience of teachers across Texas. Our state’s experience with the TAKS writing exam likewise illustrates that students who can do well on multiple-choice exams in English language arts may be unable to compose a passable expository paragraph. (“Dallas-Fort Worth students struggle with TAKS’ short-response written test,” Dallas Morning News, July 20, 2008)

Some states here in the U.S. have managed to de-emphasize multiple-choice testing for accountability. For example, Connecticut has relied more heavily on essay exams. Ironically, though, the unfunded NCLB mandate of annual testing, by making it too costly to continue the essay-exam practice, threatens to force states like Connecticut to move in exactly the wrong direction and to adopt the multiple-choice option instead.

No wonder so many employers are frustrated and keep on saying that high-school graduates aren’t prepared to apply what they’ve learned. The New Commission on The Skills of the American Workforce described the problem with the current testing regime as follows in its 2007 report, *Tough Choices or Tough Times*: “more often than not, little or nothing is done to measure many of the...qualities that we have suggested may spell the difference between success and failure for the students who will grow up to be the workers of 21st century America: creativity and innovation, facility with the use of ideas and abstractions, the self-discipline and organization needed to manage one’s work and drive it through to a successful conclusion, the ability to function well as a member of a team....”

Measuring Student Gains—Accurately

While strong state standards are fundamental and essential, other building blocks of meaningful accountability also are missing from the current system in Texas. One critical missing link is the measurement of student academic growth. The current system fails to recognize and reward improvement in student performance. The overwhelming emphasis is on reaching a fixed proficiency standard—a single benchmark of success for students on each test, a set passing rate for schools and districts—as if only that one measure of achievement matters. A new accountability system must incorporate measures of students’ academic improvement or it will forfeit credibility with educators and the public just as the current system has.

However, it won’t be easy to devise meaningful growth measures. As one scholar has put it: “(1) current methods of testing don’t measure gains very accurately; (2) some of the gains may

be attributable to factors other than the quality of a given school or teacher; and (3) we lack a firm basis for comparing gains of students of different levels of ability.” (Ballou, “Sizing Up Test Scores,” *Education Next*, Summer 2002) Current tests are not designed, for example, so that a gain of 10 points for a student well below the proficiency standard is equal to a gain of 10 points for a student at or above the standard. So a system that counts both of those score gains as evidence of the same quantum of school or teacher effectiveness will be inaccurate.

RAND Education researchers in 2004 comprehensively examined the potential of various value-added models as a basis for measuring school and teacher effectiveness and came to a similar conclusion. The RAND research team explained that no current value-added methodologies can account for the many variables beyond the current-year teacher's performance that influence student achievement: for example, controlling for student backgrounds, disentangling school and district effects from teacher effects, disentangling the effects of earlier teachers and schools from estimated teacher effects, effects of incomplete records, effects of missing records, and criteria for linking particular students to particular teachers.

They concluded that value-added methodology, given these uncertainties, is best used as just one indicator of performance that needs to be corroborated by other methods of evaluation. They said: “The research base is currently insufficient for us to recommend the use of value-added methodology for high-stakes decisions regarding employee evaluation and compensation.” (RAND Corporation, “The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness,” 2004) More recently, education researcher Audrey Amrein-Beardsley has reinforced the RAND critique of the most prominent “value-added” model propagated by statistician William Sanders, which is being used to guide development of value-added measurements at the local and state level in Texas and across the country. The RAND researchers said they could not independently corroborate the reliability of the proprietary Sanders methodology. Professor Amrein-Beardsley’s new analysis leads her to conclude likewise that the methodology is opaque and has not been independently tested enough to justify its widespread use. “The problem,” she says, “is that when these things are being sold to superintendents, they don’t know any better.” (“Scrutiny Heightens for ‘Value Added’ Research Methods,” *Education Week*, May 7, 2008)

Acknowledging the Limits of Any One Test

Even with a foundation of strong standards and even assuming we use well-founded growth measures as a performance indicator, there remain intractable problems with the use of a single state assessment to make high-stakes decisions about students, teachers, and schools. Scores on uniform state assessments may look like precise measurements. The simplicity of reducing the measurement of the complex process of teaching and learning to a single number certainly is seductive. But the appearance of clarity and simplicity is misleading.

Policy-makers need to understand just how hard it is to devise a good state assessment. State

assessments are supposed to tell us how much a student has learned, but tests are susceptible to endemic infirmities. For practical reasons, state accountability tests can only include a sampling of a wide domain of knowledge such as mathematics; if the samples of knowledge tested are inconsistent from one administration of the test to the next, there can be serious skewing of the resulting measurement of proficiency, leading to incorrect inferences about changes in student performance. Seemingly minor differences in the wording of questions, or in the wording of the wrong answers on a multiple-choice test, can also skew results. The difficulty of test items, the criteria for scoring students' work, student motivation, and the extent of drill and practice on old tests and emphasis on "eliminating the wrong answers" instead of reasoning one's way to the correct answer—all of these similarly can and do distort the picture, provided by the test, of what a student knows and can do in a particular subject area. As a result, state accountability assessments can yield seriously mistaken conclusions about student performance. (Koretz, "Measuring Up: What Educational Testing Really Tells Us," *American Educator*, Fall 2008)

The negative impact of excessive emphasis on a single state assessment goes beyond these distortions. These tests induce a reallocation of effort by educators within subject and among subjects—narrowing of the curriculum—so that students' overall mastery of their subjects suffers. Instead of leading to better teaching—which should be the goal of any worthwhile accountability system—the pressure to raise scores on one test produces negative incentives and consequences: for instance, an excessive focus on old tests and reliance on test-taking tricks. (citation: Koretz, "Measuring Up")

Even "exemplary" ratings under the state accountability system consequently can have much less substance than meets the eye. Highland Park ISD superintendent Cathy Brice, speaking from experience in a district where every school perennially receives this highest rating under the state accountability system, says: "The test measures a very narrow subset of the state curriculum standards. When we stop and do some of the more narrow curriculum to be able to max out on the state assessment, it becomes a diversion." ("HPISD Schools Exemplary Again; Outgoing superintendent pushes for less focus on TAKS," *Park Cities People*, August 8, 2008)

The implicit message of our current system to schools and teachers is that the job is done once students pass the TAKS test. In the process, schools that are not "failing" but could be better receive less attention, and students who score above the proficiency level tend to receive less attention as well. The goal should be to move students at every level of attainment to still higher levels of achievement. But instead the energy and resources needed to move the whole spectrum of student performance higher are being misdirected to serve priorities dictated by the flawed results of a flawed accountability system.

Supportive Interventions

Texas AFT believes a new system that puts standards-based reform back on track must focus on supportive interventions, not punishments, when performance as gauged by state assessments

and corroborated by other measures falls short, and must focus interventions on the schools that most need them. State law needs to stop overidentifying failure and triggering automatic sanctions, such as: removal of the principal after two years of unacceptable performance ratings, regardless of improvements shown in student achievement; removal of teachers when schools are reconstituted unless deemed to have individually improved student performance as measured by state assessments in core subjects; required arbitrary transfer of 50 percent of students and 75 percent of faculty before a school that has been shut down for repeated low ratings can be reopened; required shutdown and potential transfer to private management, even when schools and their students are actually making substantial gains in achievement.

The recent shutdown of Johnston High School in Austin illustrates much of what's wrong with current law. Although Johnston had been labeled "unacceptable" for five years, the school has been making significant progress as measured by the state accountability system. Johnston's Reading/English Language Arts scores show that ninth-grade passing rates improved over the past five years, from 35 percent to 62 percent, 10th-graders from 28 percent to 64 percent, and 11th-graders from 34 percent to 73 percent. But the new state sanctions scheme passed in 2006 gives no weight to these improvements, automatically requiring that the school be shut down.

The legislature should recognize the irrationality it has forced upon the commissioner of education, who may approve a locally developed "repurposing" plan only if 75 percent of this year's faculty and 50 percent of this year's student body are shipped elsewhere. Thanks to these arbitrary requirements, some of the very teachers responsible for Johnston's significant gains in test performance in 2007-2008, and some of the students most in need of their continued help, will be dispersed in 2008-2009.

Another case in point of the perverse effects of state law imposing sanctions for low test scores comes from Dallas ISD, where a locally devised quantitative measure of "classroom effectiveness," based largely on TAKS test scores, is now being used as the primary basis for the termination of teachers at campuses undergoing reconstitution. (State law does not require this abuse of test scores for high-stakes employment decisions, but it certainly can be said to encourage it by placing inordinate emphasis on TAKS scores in the state accountability system.) All the weaknesses of value-added methodology cited earlier seem to be in play in this Dallas ISD policy. In defense of one veteran teacher with an excellent track record who nonetheless ran afoul of this arbitrary policy, Texas AFT filed a legal challenge against its implementation by the school board, which overrode an independent hearing examiner's finding that the evidence did not support termination.

In response, the commissioner in September 2008 of education reversed the Dallas ISD termination decision. Commissioner Robert Scott noted that the state law mandating reconstitution for certain campuses does not establish cause for termination per se. He further found that evidence in the case showed other factors in the school environment, not the performance of the individual teacher, were to blame for students' low test scores. Other pending

cases will tell us whether this state ruling was an outlier or a sign that the tide is turning toward more common sense in the use of students' test scores for high-stakes employment decisions.

Texas AFT strongly supports the intelligent redesign of weak instructional programs, backed up by the deployment of needed resources, when neighborhood schools are persistently low-performing. But the development of local redesign plans for particular campuses should be free of arbitrary criteria such as mandatory removal of fixed percentages of staff and students—let alone the arbitrary termination of employment of teachers at such campuses.

Appropriate Testing of Students with Disabilities

Fair assessment of students also demands that we restore common sense to the testing of students with disabilities, too many of whom, per state rules responsive to NCLB, are being required to receive on-grade-level instruction and take on-grade-level TAKS tests that are flatly inappropriate under the Individuals with Disabilities Education Act. For a student in special education, the Individualized Education Program (IEP) should govern instruction and testing, and the regular-education teacher should be part of the IEP decision-making team.

In its zeal for accountability, the federal government has lost sight of how to help special-education students, creating severe contradictions between federal testing requirements under the No Child Left Behind Act and the appropriate-education requirements of the federal Individuals With Disabilities Education Act. Many students in special education can handle on-grade-level work, but the problem is the arbitrary requirement that 97 percent (1 percent allowed to take TAKS-ALT, 2 percent allowed to take TAKS-M) of special-education students must be tested on grade level, regardless of what is appropriate for each individual student. Otherwise, their passing scores cannot be counted as proficient for purposes of meeting the federal requirement of adequate yearly progress (AYP).

This requirement is a mandate for educational malpractice. As one special-education teacher noted, she is forced to test her students on Algebra I, even though these students are operating on a second-grade level and have not mastered basic math concepts of addition and subtraction. Compounding this absurdity, these same students will likely be required to attend summer school to study Algebra I, instead of receiving what they really need first—intensive instruction to help them learn the foundation concepts they have yet to master. This approach not only demoralizes the students, it also halts their educational progress by attempting to force-feed them advanced concepts for which they are simply not ready at their current stage of development.

It makes no sense to treat these children or their schools as failures for not meeting these arbitrary standards that assume nearly every student is on grade level, regardless of disability. The proper measure of a special education student's progress should be his or her attainment of Individualized Education Program (IEP) goals.

Before passage of NCLB, the state of Texas already had developed appropriate testing requirements that met students' needs, raised expectations for students with disabilities, enjoyed broad support among special-education advocates, and minimized any risk of over-identification of special-education students. While we agitate for corrective action in Congress, the state of Texas should do everything in its power to resist the federal folly that has been superimposed on our state accountability system under NCLB. Specifically, the Texas Education Agency should make not a single further move to align state special-education accountability requirements with current, arbitrary NCLB mandates that are not in the best interests of our special-education students and are highly likely to be revised.

Appropriate Testing of English Language Learners

The current accountability system also creates pressure for the inappropriate testing of ELLs. Testing kids in English who don't speak English isn't going to tell you much except that they don't speak English. Yet our accountability system puts too many students in exactly this box.

Research clearly demonstrates that, in reading for example, learning to read in the home language promotes reading achievement in the second language. Five independent meta-analyses based on experimental studies all have reached this same conclusion. The likely explanation for why instruction in the native language actually promotes English proficiency is the "transfer" phenomenon. That is, if you learn something in one language, you either already know it in (i.e., transfer it to) another language or can more easily learn it in another language. (Goldenberg, "Teaching English Language Learners: What the Research Does—and Does Not—Say," *American Educator*, Summer 2008)

Inappropriate testing of ELLs in English therefore likely leads to underestimated abilities and underestimated accomplishments of our schools. One Texas teacher in a letter to Texas AFT put it this way: "I am writing this on behalf of my current class of fourth-grade ESL students, who are transitioning this year from the bilingual program...My problem is that, in a few months, these children will be required by the state to take the TAKS test in English, starting in February with the Writing portion, and finishing with Math and Reading in April. My students are as bright and motivated as any others, more so than many; however, they simply have not yet acquired the English skills necessary to pass these exams."

The teacher went on to write that she and her colleagues wanted the state to consider giving these children an extra year or two to make the transition into English-only testing. "This would give them a chance to build some confidence and develop some of the skills required to succeed in English," she wrote. Alternatively, she said, "If they must take the TAKS, why not use it as a skill indicator and needs assessment rather than a pass/fail threat?" Why not, indeed.

An AFT Alternative to the Current System

Beyond TAKS (and NCLB): Putting Texas School Accountability Back on Track (Discussion Draft, October 3, 2008)

Absorbing the lessons of all the weaknesses and perverse consequences of the current test-driven accountability system, the goal for policy-makers should be to create an alternative system of accountability that gauges a broad array of factors contributing to learning. The new system should take into account and share with stakeholders not just state assessment scores but also other factors known to affect student achievement. These should include facility conditions, classroom safety, measures of support and impediments in the community, and availability of essential resources.

The United Federation of Teachers, AFT's affiliate in the New York City school system, has developed a draft model of such a comprehensive accountability scorecard that Texas AFT commends to the attention of Texas policy-makers. UFT's proposal first provides a baseline inventory ("school portrait") of critical factors affecting achievement at each campus, including per-pupil budget, percentages of high-need students (economically disadvantaged, special education, ELL), percentage of faculty teaching out of field, class sizes, condition of facilities, and availability of instructional materials.

The UFT model then provides a letter-grade assessment of each campus in each of four dimensions:

- academic achievement;
- safe learning environment;
- teamwork to promote achievement;
- and resources provided by state and district to meet students' needs (e.g., funding, quality professional development).

The UFT proposal looks at each of these dimensions not just for the current year but over a three-year span as well, and it compares a school's status in each dimension with that of other similar schools, as measured by their percentage of high-need students (Title I eligible, special education, ELL). For the sake of further comparison, a district-wide average is noted for each of the four dimensions of accountability.

Academic Achievement

UFT's model uses multiple measures of academic achievement. Passing rates on standardized tests are just one sub-part of the UFT gauge of campus achievement. Also a factor at the high-school level, for example, is the graduation rate, broken out by subgroups. While these factors account for half of a school's grade on academic achievement, another 35 percent is based on measures of year-to-year progress. (Such progress presumably could be measured by progress of individual students from year to year or performance of successive student cohorts at the same grade level from year to year.) The remaining 15 percent is based on the scope of the curriculum.

Safe Learning Environment

Also graded under UFT’s model is the safety of the school learning environment—a crucial prerequisite for effective learning and instruction. Partly measured by the incidence of major and minor crimes on campus, performance in this dimension also is measured by clarity of the campus code of conduct, dissemination of these rules to all stakeholders, and consistent enforcement.

Teamwork to Promote Achievement

Teamwork to promote achievement, the third basic dimension of campus performance under the UFT plan, encompasses the quality and quantity of collaboration among members of the educational team, the engagement of educators with parents, parental involvement, student commitment (e.g., attendance), staff absenteeism, and staff turnover.

Resources

The school’s fourth and final letter grade under the UFT scorecard model of accountability is really a grade for the “management” of the school—those responsible for providing the resources needed to get the job done. This dimension of accountability reflects:

- resources provided (e.g., budget per pupil, quality professional development, mentoring, facilities, materials);
- administrative support (class size, provision of counseling/social work services, provision of special-education and ELL services, transparency of management decision-making);
- and quality of the curriculum (e.g., provision not only of quality curriculum in core courses but also of quality curriculum in foreign languages, physical education, career and technology courses, and music and visual arts).

A Spur to School Improvement

The detailed information on all four dimensions of school performance becomes the basis for a school improvement plan. That plan is forged through internal and external evaluation of the school’s performance across all four dimensions, with a school leadership team doing the internal evaluation and an independent quality review team providing the external evaluation.

Texas accountability law already requires a campus inventory addressing many of the same factors identified in the UFT model, to be conducted by a campus intervention team after two years of unacceptable ratings under the state accountability scheme. But the provision (Texas Education Code Section 39.1323) lacks a key ingredient: an honest assessment of the adequacy of resources provided to meet student needs—the critical fourth dimension of the UFT model.

The Texas provision authorizes a campus intervention team merely to suggest “reallocation of resources,” not additional resources, to address campus performance shortfalls. And of course the measure of performance is one-dimensional in the extreme, addressing primarily passing rates on state standardized tests.

A Balance of Qualitative and Quantitative Accountability Criteria

The UFT model, which demands extensive, detailed qualitative as well as quantitative evaluation of school and system-wide effectiveness, implies a big investment in outside expertise as well as improved tests that measure higher-order achievement. To illustrate what might be involved in a statewide application of this model, consider the notion of a “statewide review board that would function independently of the state’s department of education,” as sketched by education scholar Helen Ladd, in a recent essay on “Rethinking the Way We Hold Schools Accountable.” (Education Week, January 23, 2008) In Professor Ladd’s vision of a more balanced system, states would still use test scores to hold schools accountable for “realistically obtainable gains in student performance in core subjects.” But states would supplement their reliance on test scores with “a new system of school inspections designed to improve practice and to encourage schools to pursue many of the other outcomes demanded of a good education system.” (Ladd cites New Zealand’s system of school inspections as the inspiration for her proposal.)

Ladd’s proposed statewide review board “would send small teams of professionals to make periodic visits to each school—perhaps one visit every two or three years—with each visit preceded by an internal self-study. The review panel would then write a report on each school that, along with the school’s response, would be made public. Though the report would include a summary of the school’s success, or lack thereof, in raising student achievement in the core subjects, it would evaluate the school on a far broader set of outcomes than student test scores alone....the review panel would look closely at the policies and systems that schools put in place to promote those outcomes. The panel itself would not be in the business of providing assistance or support to the school, since doing so would interfere with its ability to be objective”—including the ability to be objective about the adequacy of resources provided to the school to get the job done. “District and state policymakers would be expected to make use of these reports in allocating resources, providing technical assistance, or otherwise making sure schools have access to the capacity they need.”

Such a system would not come cheap. But if positive impact on student outcomes is the ultimate point of the accountability system, this balanced model of accountability offers a better chance of success than the current test-based system.

Ladd’s point about targeting “realistically obtainable gains” in student performance deserves extra emphasis. The standard for progress under a new accountability system needs to be grounded in empirical evidence of what is possible. Thus, a standard might be based on the actual progress recorded over several years by the top quartile of schools with substantial

populations of economically disadvantaged students. Another option would be a standard based on average growth across the state.

Implications for Use of Test Scores in Professional Evaluation

The many failings of state assessment instruments recited previously support the conclusion that test scores should not be used as the sole or principal basis for high-stakes employment decisions: evaluation, compensation, termination. Yet the Texas legislature and some individual school districts have plunged headlong into the funding of experiments with incentive schemes tied in large measure to test scores.

Texas AFT takes the view that no such compensation policy should be allowed except under a policy developed with teachers involved as full partners and approved by those affected. Such experiments should have at their center a focus on improving teaching practice, not sorting and ranking teachers according to flawed value-added measures based on test scores. It should be clearly understood from the outset that even the best of these programs are indeed experiments, ventures into uncharted territory, entered into hopefully but with no assurance that they will work to lift up student achievement.

A Look at Current Experiments

We encourage policy-makers to look carefully at the lessons to be drawn from the design of two of the most frequently cited examples of model “performance pay” schemes, which actually point to the importance of many factors other than incentive pay as drivers of educational improvement.

The Teacher Advancement Program developed by the Milken Foundation, for instance, stresses much more than measurement of student test scores as a mechanism to drive achievement. For example, TAP emphasizes: creation of opportunities for teachers to advance their professional career and compensation without leaving the classroom, by serving as master teachers and as mentors; assurance of time during the regular school day for teachers, master teachers, and mentors to meet and plan improved instruction to address student needs; intensive evaluation of each teacher’s work, four to six times a year, with reference to multiple criteria, to inform and support improved practice; extra pay based on the extra demands placed on the individual teacher as well as on student achievement, with all teachers eligible for the extra pay; overwhelming support for the experiment from the teachers subject to it before the experiment commences; and increased resources of roughly \$400 per pupil to finance the experiment. (“Teacher Advancement Program,” National Institute for Excellence in Teaching, 2006; additional citations forthcoming)

The Benwood performance-pay initiative in the public schools of Chattanooga, Tennessee, also illustrates the importance of factors other than bonus pay in improving educational performance. The key concept in this experiment has been to offer extra pay to induce veteran teachers,

selected based on the track record of their students on standardized tests, to accept assignments at low-performing schools. But a key finding of the Benwood experiment to date is that improved professional development for teachers on those low-performing campuses has contributed at least as much to improved test scores.

The improvement in professional development, as in the TAP experiments around the country, is just one part of a significant infusion of new resources invested to lift up the performance of low-performing Chattanooga schools. As one favorable commentary on the Benwood initiative stated, “the district invested heavily in mentoring programs to train teachers, in additional staff to support curriculum and instruction, and in stronger and more collaborative leadership at the school level. At the same time, the Benwood initiative was buoyed by better labor-management relations and a host of other reform efforts at the district level.” (Silva, “The Benwood Plan: A Lesson in Comprehensive Teacher Reform,” Education Sector, April 2008) In one sense, the commentator concludes, the lesson of Benwood is sobering—other districts are not going to be able to replicate Benwood’s success just by implementing a performance-pay plan. However, “the steady, marked increase in the effectiveness of Benwood teachers suggests that teacher effectiveness isn’t fixed.” The Benwood experiment shows that “many teachers who are currently struggling to help disadvantaged students can do much better” with the help of resources and tools to improve as professionals and with community support and recognition.

A Texas experiment along similar lines is under way in Austin ISD. There Texas AFT’s local affiliate, Education Austin, has negotiated a pilot plan for “strategic compensation” at nine campuses, with the approval of a super-majority of the faculty at each campus. Teachers are invited to come up with appropriate measures by which to judge their effectiveness. These measures are deliberately not TAKS-based. The emphasis is on marshaling necessary resources and building up the capacity of individual teachers and their colleagues to provide effective instruction. The stated aim is to test a new model of compensating teachers that enables practitioners to “augment their base salary while also engaging in new practices to increase their professional skills and improve student learning.”

What may we infer from these examples? They are still experiments, not proven programs, but they share promising features that would be reinforced under our recommended alternative to the current accountability regime: emphasis on support, not sanctions; recognition of the multiple factors that affect student achievement; a focus on fine-grained analysis and resulting action to improve instruction in every classroom and to improve learning for all students.

Lessons from Business Practices

Self-described “reformers” ardently advocate “performance pay,” evaluations, and even terminations based on the results of flawed state assessments of the students in an individual teacher’s classroom. Often these advocates say their aim is to make our schools run “like businesses.” Ironically, much of the business world has moved away from this approach to

employee pay and evaluation. As Richard Rothstein has written: “Most private-sector jobs, like teaching, include a composite of easily measured and less-easily measured responsibilities....A widespread business reform in recent decades has been totally quality management, inspired by W. Edwards Deming, who warned that businesses seeking to improve quality and thus long-term performance should eliminate work standards (quotas), eliminate management by numbers and abolish merit ratings and management by objective because all of these encourage employees to focus on short-term results.” (Rothstein, “The Corruption of School Accountability,” *The School Administrator*, June 2008)

Rothstein also cites “the balanced scorecard” as “a corporate accountability tool that has grown more recently in popularity” because “business management theorists concluded that quantifiable short-term financial results were not an accurate guide to future profitability. Firms’ goals were too complex to be reduced to a few quantifiable measures because predicting future performance relies not only on past financial success, but on subjective judgments of product quality, employee motivation, internal corporate cohesion and customer satisfaction and loyalty.”

It is highly instructive to note that, even as test-score fetishism in education reached new heights under the No Child Left Behind Act, the federal Department of Commerce continued to use a “balanced scorecard” approach in administering the Malcolm Baldrige National Quality Awards for exemplary institutions in manufacturing and other business sectors. “Numerical performance indicators play only a small role in award decisions,” Rothstein observes. Furthermore, “the Baldrige award program and its principles were extended to health and education institutions in 1999. For school districts, only 100 of 1,000 points are for student learning outcomes, with other points awarded for subjectively evaluated measures, such as ‘how senior leaders’ personal actions reflect a commitment to the organization’s values.” One more irony noted by Rothstein: The most recent award for elementary and secondary education was given in 2005 to the Jenks, Oklahoma, school district, based on factors reaching far beyond students’ test scores, such as low teacher turnover, an international-exchange program, and the enlistment of residents to mentor pre-kindergartners and kindergartners. Yet in 2006 the Jenks district was rated substandard under NCLB because students failed to meet federal AYP targets for reading test scores.

Control Over Instructional Time

The current high-stakes TAKS test drives administrators to mandate intensive test preparation in quest of higher test scores. School districts are pushed in this direction by state performance reviews explicitly as well as by the intrinsic nature of a test-driven accountability system. Districts procure test-prep materials from test publishers and expect these materials to be used. Principals are under the gun from both the state (with automatic removal and reassignment after two years of “unacceptable ratings” at their campus) and from district central offices.

In the face of these multiple perverse incentives to devote inordinate time to test preparation narrowly conceived, restoring authority over instructional time to the individual teacher in the classroom becomes a sine qua non of an alternative accountability system, as far as Texas AFT is concerned. Teachers should be empowered to decide, based on the particular needs of their students, whether or when class time should be diverted from regular instruction to preparation for state assessments.

As documented earlier in this report, it is generally agreed that the professional skill of the individual classroom teacher is the most important school-related factor affecting student achievement. A reformed accountability system must allow teachers to use their professional judgment and skill to strike the proper balance between testing and teaching and learning in their classrooms.

The Texas legislature last year in SB 1031 tried to limit district-mandated tests to 10 percent of annual instructional days. That provision is now codified as Education Code Section 39.0262. However, school districts around the state have continued to claim that much of their testing is not district-mandated but rather discretionary at the campus level. In Austin ISD, for example, our members consequently reported in February of this year that they had spent three and a half of the previous seven weeks giving tests instead of teaching. On one campus, the English department met to develop new curriculum ideas to address the school's "unacceptable" test scores, yet two months into the semester the teachers still lacked the time to implement the promising curriculum changes they had identified. Why did they lack the time? Classroom time already was preempted by the next rounds of required testing.

The statutory cap on local testing is a good idea, but in many districts the provision clearly is not working as intended to reduce the proportion of instructional days spent on testing. We call on the legislature to strengthen this statutory cap by amending it to cover all locally mandated testing, not just explicitly "district-mandated" testing—with a proviso that the limit of 10 percent of instructional days does not apply to testing approved or initiated by the individual classroom teacher.

The best safeguard of an appropriate balance between testing and instruction for students is to let the individual classroom teacher make a professional judgment as to when and if testing in excess of 10 percent of instructional days is appropriate. Such an amendment of the testing cap would put local control in the hands of the individual professional educator who is in the best position to judge day by day whether instructional time should be spent on additional testing. Of course, the ultimate goal we have in mind is an alternative system that puts testing back in its proper place and reverses the current overwhelming pressure for "drill and kill."

Factors Outside the Schoolhouse

Putting testing back in its proper place means a new accountability system ultimately must acknowledge that schools are part of the solution but cannot alone solve performance woes that trace in large measure directly to socioeconomic disadvantage. If we are honest with all stakeholders in the public schools about what it takes to lift up student achievement to meet international competition, we need to face and tackle those socioeconomic disadvantages, with universal access to early childhood education, timely health care for every child, education for parents of our schoolchildren, living-wage and affordable-housing programs, and more.

Recently AFT President Randi Weingarten presented a vision of the role schools could play in this expanded arena of public accountability to our students for the help they need to succeed. Weingarten envisioned “community schools” that bring together “under one roof all the services and activities they and their families need.” (Weingarten, AFT Convention Speech, July 14, 2008) She invited us to imagine schools open all day, offering after-school and evening recreational activities as well as homework assistance, with the high schools allowing students to sign up for morning, afternoon, or evening classes. She pictured schools that could offer child care and dental, medical, and counseling services, along with English-language instruction for parents, GED programs, and other coordinated assistance.

Along with this outward orientation to address problems arising outside the schoolhouse, Weingarten envisioned the changes within schools that a broader notion of accountability would help to foster. Mentoring for new educators and peer coaching for those who are struggling, common planning time across disciplines, collaborative relations with administrators, compensation models that support great teachers and keep them teaching all are parts of that vision.

Along similar lines, a distinguished group of national education scholars has issued a statement calling for “A Broader, Bolder Approach to Education” to replace current policies “crafted around the expectation that schools alone can offset the full impact of low socioeconomic status on learning.” (Ladd et al., “A Broader, Bolder Approach to Education,” June 10, 2008)

The scholars don’t dispute that schools can ameliorate some of the impact of social and economic disadvantage. However, they say, “Despite the impressive academic gains registered by some schools serving disadvantaged students, there is no evidence that school improvement strategies by *themselves* can close these gaps in a substantial, consistent, and sustainable manner.”

They add that there is strong evidence for the effectiveness of a combination of policies “aimed directly at education-related social and economic disadvantages” that “can improve school performance and student achievement.” The scholars further assert: “The persistent failure of policy makers to act on that evidence—in tandem with a school-improvement agenda—is a major reason why the association between social and economic disadvantage and low student

achievement remains so strong.”

Hence these scholars advocate an expanded concept of education, encompassing learning that occurs outside of formal schooling, in early childhood and pre-school programs, after-school and summer programs, and via programs that develop parents’ capacity to support their children’s education. They also stress more than basic academic skills, addressing the development of the student as a whole person, including physical health, character, social development, and non-academic skills—and particularly the new knowledge, skills and creativity needed to participate in a competitive global environment.

The authors of the “Broader, Bolder Approach” accordingly call for increased investment in developmentally appropriate and high-quality early childhood, pre-school, and kindergarten education; for school-improvement efforts including small class sizes for disadvantaged children, measures to attract high-quality teachers to work in hard-to-staff schools, better alignment of pre-school, elementary, secondary, and higher education; and for increased investment in health services for pregnant women and preventive and routine pediatric, dental, and optometric care for all infants, toddlers, and schoolchildren.

“By and large,” they say, “low-income students learn as rapidly as more-privileged peers during the hours spent in school. Where they lose ground, though is in their lack of participation in learning activities during after-school hours and summer vacations.” Therefore, the scholars say, policy-makers should invest more in longer school days, after-school and summer programs, and school-to-work programs with demonstrated track records.

Where does an alternative school accountability system fit into this picture? The authors of the “Broader, Bolder Approach” answer: “The public has a right to hold schools accountable for raising student achievement. However, test scores alone cannot describe a school’s contribution to the full range of student outcomes. New accountability systems should combine appropriate qualitative and quantitative methods, and they will be considerably more expensive than the flawed accountability systems currently in use by the federal and state governments.”

Naysayers will argue that we are looking for ways to adjust the accountability system to take the pressure off of schools, to excuse low performance. On the contrary, our alternative accountability system would expose to full public scrutiny and put more pressure on all the players in the public arena—from the schoolhouse to district headquarters to Austin to Washington—to do their part to deliver on the promise of school improvement. Instead of the current, top-down model of accountability, the alternative pictured here would make policy-makers and schools alike more accountable to the parents and taxpayers at the grass roots who want neighborhood schools that serve their students well.

If, however, the naysayers’ complaint is simply that we would de-emphasize their preferred accountability metric of standardized state test scores, then we invite them to consider this fact:

No country whose students outperform our own even has a test-based accountability system, let alone one like ours that only imprecisely measures relatively low-level skills across a narrow arc of the curriculum.

Independent Review of Any New System

Lawmakers need to recognize that nobody has a perfect model of accountability (and we certainly make no exception for ourselves). Too much remains uncertain about some of the fundamental ideas involved in measuring accountability. For this reason we would add one indispensable further requirement to our criteria for any new accountability system that would replace NCLB and TAKS. That criterion is a requirement of external vetting by independent experts, so that we maintain due respect for the boundary between what we know to be true and what we wish were true regarding our ability to measure the performance of students, teachers, and schools.

Conclusion: A Vision of Accountability Based on Classroom Realities

Our vision of a reformed accountability system is based on classroom realities. Schoolchildren are not widgets, and teachers are not cogs in a machine. The current version of accountability has degenerated into a system of what one business writer calls “the folly of accountabilism”—a fetishism of test scores, based on the spurious precision of the numbers that purportedly define success or failure for students, teachers, and schools. “Accountabilism bureaucratizes and atomizes responsibility. While claiming to increase individual responsibility, it drives out human judgment....Accountabilism turns complex systems into merely complicated systems, sacrificing innovation and adaptability. When a sign-off is required for every step in the work flow, those closest to a process lack the leeway to optimize or rectify it.” (Weinberger, Harvard Business Review, February 2007))

Texas AFT’s alternative to accountabilism—the multi-factor, diagnostic analysis of the efforts of students, teachers, their schools, and their community—is grounded in an understanding of the central, professional role of the individual teacher confronting the myriad individual needs of students in the classroom. Nobody has said it better than the late Al Shanker, who as president of the American Federation of Teachers was a national leader in standards-based reform:

“Top-down management does not work. Neither does top-down reform. We cannot help Johnny overcome his reading problem by turning to page 234 of a state regulation. The people who wrote those regulations are not qualified teachers, nor have they spent six months in the classroom observing Johnny and trying out and discarding four different approaches to solving his particular difficulty. The fifth approach—the one that may work—is not to be found in a state law or a school district’s administrative directive. It can only come from the mind and hands of a

creative and sensitive teacher.

“Teaching, like medicine, cannot operate by remote control. There is no formula that fits all children. The only treatment that works is one that is constantly adjusted and fitted and fine-tuned by the people on the scene. Intelligent change has its best hope in teachers because nobody knows better than teachers what is going on in schools.” (Shanker, “Our Profession, Our Schools: The Case for Fundamental Reform,” *American Educator*, Fall 1986)

Following in the tradition of Al Shanker, current AFT President Randi Weingarten has summed up our entire argument for a new accountability system in one sentence: “What we need is accountability that is meant to fix schools, not to fix blame.” (Weingarten, 2008)