



**Written Testimony to the Senate
Committee on Health, Education,
Labor and Pensions (HELP) on
“Fixing No Child Left Behind:
Testing and Accountability”**

January 30, 2015

The Association of Texas Professional Educators (ATPE) is the preeminent educator association in Texas and makes a positive difference in the lives of educators and schoolchildren. ATPE is a member-owned, member-governed professional association with more than 100,000 members, making it the leading educators’ association in Texas and the largest independent association for public school educators in the nation. ATPE submits this testimony in support of limiting the negative impact on K-12 public education by the overuse of high-stakes standardized tests as required by the Elementary and Secondary Education Act (ESEA), also known as the No Child Left Behind Act (NCLB).

NCLB and the Texas testing regime upon which it was modeled were originally designed to identify education gaps, encourage the closing of those gaps, and measure progress toward closure. However, despite the use of standardized testing skyrocketing nationally in 2002 with the enactment of NCLB and its mandate of annual testing in all 50 states, the emphasis on testing has yielded little learning progress but caused significant harm, according to a nine-year study conducted by the National Research Council (NRC)¹, which concluded that testing has highlighted gaps but has done little to nothing to impact those gaps directly. Paradoxically, the fixation on testing has prohibited focusing on finding solutions to the problems testing originally highlighted—it’s as if we as a country have focused on diagnosing the patient over and over again without ever following up the diagnosis with a significant course of treatment.

ATPE and other stakeholders in Texas believe the current testing regime driven by NCLB is ineffective and even harmful for students, insofar as high-stakes testing leads to a narrowing of the curriculum, hinders differentiated instruction, and leads to numerous problems, such as student anxiety and the potential for cheating. We offer recommendations for Congress to give states more flexibility to innovate and choose assessment methodologies that better suit the needs of their students, parents, and educators.

¹ Elliott, S. and Hout, M., *Incentives and Test-Based Accountability in Education* (2011).

The current testing regime is ineffective

1. High-stakes testing is an unreliable measure of student performance

The Brookings Institution published a study in 2001 finding that 50 to 80 percent of year-to-year test score improvements were temporary and “caused by fluctuations that had nothing to do with long-term changes in learning.”² The most comprehensive national research, conducted by the National Research Council (NRC), indicates that test-based incentives increase teaching to the test and are an inflated and inaccurate representation of student knowledge.³ The NRC concluded in particular that tests required for high school graduation have not increased student achievement but have raised the dropout rate an average of two percent.⁴

2. High-stakes testing does not improve student achievement

As a result of our test-driven culture, US students have shown little to no improvement nationally and have either slipped or remained stagnant in global performance. We can look at performance by US students on the Programme for International Student Assessment (PISA), a triennial evaluation of world education systems that tests 15-year-olds in approximately 70 participating economies, as an indicator. In 2012, the United States performed below average in mathematics and was ranked 27 among the 34 Organization for Economic Co-operation and Development (OECD) countries—a forum of participating countries that exists to allow for the sharing of best practices among governments and economies, which is also responsible for PISA. More than one-fourth of US students failed to reach the PISA baseline level (Level 2) of mathematics proficiency, a statistic that has not changed since 2003, and the United States has a below-average share of top performers in mathematics, with only 2% of US students reaching the highest level (Level 6). Meanwhile, although students performed near the average in reading (ranked 17) and science (ranked 20), key findings from the PISA results in 2012 state there has been no significant change in performance by US students since 2002⁵—both when PISA was first administered and when standardized testing grew substantially in the United States.

On May 26, 2011, an NRC report found that there is no evidence that test-based incentive programs are working. “Despite using them for several decades, policymakers and educators do not yet know how to use test-based incentives to consistently generate positive effects on achievement and to improve

² Lynn Olson, “Study Questions Reliability of Single-Year Test-Score Gains,” *Education Week*, May 23, 2001.

³ “FairTest’s Testimony to the Rhode Island Legislature on Graduation Exams,” www.FairTest.org, Feb 26, 2014.

⁴ *Id.*, citing Hout & Elliott, 2011.

⁵ Organisation for Economic Co-operation and Development (OECD), “Programme for International Student Assessment (PISA) Results from PISA 2012” <http://www.oecd.org/pisa/keyfindings/PISA-2012-results-US.pdf>

education.”⁶ Significantly, the United States is one of only three of the 34 OECD countries that rely not only on national examinations, like the NAEP, but also on other, non-national types of assessments in elementary and secondary education, like the statewide assessments called for in NCLB.

In addition, urban graduation trends have reversed since NCLB was passed. Studies show that from 1996 until 2002, 68 of the 100 largest urban districts had rising graduation rates.⁷ But, from 2002 until 2006, 73 of the 100 largest districts experienced declining graduation rates. The data suggest that instead of offering a pathway to success and increased student graduations, NCLB has had a negative impact on students living in urban areas.⁸

3. High-stakes testing does not measure many important educational qualities

Late education researcher Gerald W. Bracey, PhD, said that standardized tests do not measure “creativity, critical thinking, resilience, motivation, persistence, curiosity, endurance, reliability, enthusiasm, empathy, self-awareness, self-discipline, leadership, civic-mindedness, courage, compassion, resourcefulness, sense of beauty, sense of wonder, honesty, or integrity.”⁹

In 1958, the Torrance Tests of Creative Thinking (TTCT) were developed by E. Paul Torrance to measure student creativity. Torrance and a colleague tracked kids for 40 and 50 years and found the TTCT to be a three-time stronger correlation to creative achievement than IQ.¹⁰ In May 2010, the TTCT were re-normed by College of William and Mary researcher Kyung Hee Kim, who discovered a pattern of decline in scores since 1990.¹¹ Many have said that a possible reason for a decline in US student creativity is that “standardized tests may be bad for kids.”¹²

According to the PISA results in 2012, even higher cognitive thinking can be a problem for US students. “Students in the United States have particular weaknesses in performing mathematics tasks with higher cognitive demands, such as taking real-world situations, translating them into mathematical terms, and interpreting mathematical aspects in real-world problems.”¹³

4. High-stakes testing does not align with college standards

⁶ Committee on Incentives and Test-Based Accountability in Public Education at the National Research Council, *Incentives and Test-Based Accountability in Education*, www.nap.edu, 2011.

⁷ FairTest, www.FairTest.org, citing Advancement Project (2010) (downloaded Aug 2014).

⁸ *Id.*

⁹ Valerie Strauss, “The Myths of Standardized Testing,” *Washington Post*, Apr. 15, 2011.

¹⁰ “The Creativity Crisis”: A Conversation with Nurture Shock Author Ashley Merryman, Jul 21, 2010, <http://www.learningfirst.org/creativity-crisis-conversation-nurture-shock-author-ashley-merryman>.

¹¹ *Id.* “The Creativity Crisis” 2010

¹² *Id.*

¹³ *Id.* Organisation for Economic Co-operation and Development (OECD) 2012

Many states, including Texas, have worker shortages in the skilled trades, and improvements on state tests aren't reflected in college entrance exams such as the SAT or ACT. In fact, average SAT reading scores in Texas have declined eight points in the last decade. A recent review by the National Academy of Sciences shows that high-stakes testing is not improving academic achievement and may do more harm than good. While there is a debate about how to calculate dropout rates, most agree that Texas has a serious problem with high school dropouts.¹⁴

The current testing regime is harmful

1. Excessive spending on testing diverts valuable resources

The enactment of state and federal accountability laws, including NCLB, resulted in substantially increased spending by states on testing. The testing explosion has necessitated more government spending on developing, field-testing, and administering tests; buying test prep materials; funding remediation programs for students who fail the tests; administering pre-test “benchmark” assessments at the district level; training staff and hiring additional staff for administration and evaluation of the tests; and so much more. Over a two-year period, the Texas Education Agency (TEA) spends nearly half a billion dollars just on its contract with test vendor Pearson. This does not include all of the additional spending on standardized testing by Texas school districts. The resources spent on testing in Texas and throughout the country are excessive and are spent in lieu of spending on population growth needs, smaller class sizes, improving teacher quality and pay, increased support for struggling schools, and more.

2. High-stakes testing is discriminatory against non-English speaking, low socioeconomic, and special needs students

Students from low-income families and students of color suffer from the high-stakes testing regime at a higher rate than other students. Decades of research has demonstrated that black, Latino, and Native American students, as well as students from some Asian groups, experience problems with high-stakes testing. For many of these students, there is no pathway to success under our current test-driven system, and as a result, they are most definitely being “left behind.”

The problems for these subpopulations include disproportionate failure to pass a one-size-fits-all high-stakes test tied to graduation, which provides no social or educational benefit and does not improve college or employment readiness¹⁵; a higher likelihood of low test scores and, therefore, grade retention, which causes students in these groups to progress more slowly, suffer significant loss of self-esteem, and

¹⁴ The American Conservative, *Texas vs. No Child Left Behind* (January 23, 2013).

¹⁵ “Racial Justice and Standardized Educational Testing,” www.FairTest.org (downloaded May 2014).

become less likely to graduate¹⁶; creation of a bias against all testing that results in lower scores on college entrance tests (SAT and ACT), contributing to the racial gap in college enrollment and completion¹⁷; an inaccurate measure of academic ability and employment readiness for English language learners (ELLs) and students with disabilities, who are often required to complete required testing in English before they have mastered the language¹⁸; the chance of being suspended, expelled, “counseled out,” or otherwise removed from school due to low test scores, in an effort to boost school achievement results and escape test-based sanctions mandated by NCLB¹⁹; and disproportionate misplacement of students of color in special education programs based on test results.²⁰

The one-size-fits-all standards and curriculum are normed on white, middle to upper class experiences and cultural practices, and thereby overlook and exclude entire populations of students.²¹ Data from the 2012 PISA assessment found that socioeconomic background has a significant impact on student performance, and disadvantaged students show less engagement, drive, motivation, and self-beliefs.²² Likewise, although students in special education are required by federal and state law to have an Individualized Education Plan (IEP), which can allow for a significantly modified curriculum and instruction for these students, they are measured by the same standardized yardstick as students without an IEP. Standardized tests frequently do not align with many of the goals and objectives listed in a student’s IEP, and the student will receive few accommodations, if any, that are provided to him in normal, non-test settings as part of his IEP.²³

Similarly, much of the above can be translated to the broader student body perspective; the one-size-fits-all model of standardized testing inaccurately measures ALL students on the same yardstick by assuming they are all on the same path to success. Tests do not vary in content or expectations of students based on differing goals, and they assume state or federal bureaucrats know what is best for students hailing from extremely diverse settings (Houston versus a rural town, for instance). The one-size-fits-all

¹⁶ Neill, M., “A Child is Not a Test Score: Assessment as a Civil Rights Issue,” *Root & Branch*, Vol. II, Issue 2, Fall 2009.

¹⁷ “Racial Justice and Standardized Educational Testing,” www.FairTest.org (downloaded May 2014).

¹⁸ National Council of Churches Committee on Public Education and Literacy, “[Ten Moral Concerns in the Implementation of the No Child Left Behind Act](#),” www.nccusa.org (accessed June 21, 2011).

¹⁹ “Racial Justice and Standardized Educational Testing,” www.FairTest.org (downloaded May 2014), citing Advancement Project, et al. 2010, *Federal policy, ESEA Reauthorization, and the School-to-Prison Pipeline*, and Neill, M., Guisbond, L. & Schaeffer, B. 2004, *Failing Our Children*, www.FairTest.org.

²⁰ “Racial Justice and Standardized Educational Testing,” www.FairTest.org (downloaded May 2014), citing Neill, et al. 2009, *Op. cit.* FairTest Examiner, Jan 2007, “NCLB Reforms for Disabled Students.”

²¹ “Racial Justice and Standardized Educational Testing,” www.FairTest.org (downloaded May 2014), citing Neill, et al. 2004, *Op. cit.* FairTest Examiner, Jan 2007, “NCLB Reforms for Disabled Students.”

²² PISA – US, p. 1.

²³ Katherine McKnight, PhD, “Opting Out of NCLB Testing,” www.katherinemcknight.com, Mar. 25, 2011.

testing model impacts all students by failing to measure success in relation to an individual student's goals or intended path. The result can be many of the unintended consequences previously mentioned for minority and special needs populations, including an increased likelihood to drop out because students don't see education as relevant to them.²⁴

3. High-stakes testing narrows curriculum

High-stakes testing encourages a narrowed curriculum that undermines rather than improves education.²⁵ FairTest has found that untested subjects are ignored while tested subjects are narrowed to the tested material. Some schools allocate more than a quarter of the year's instruction time to test preparation, and students learn merely to identify correct answers to multiple choice questions instead of learning to apply knowledge to problems to find solutions.²⁶

A 2007 national study by the Center on Education Policy found that since 2001, 44 percent of school districts had reduced the time spent on science, social studies, and the arts by an average of 145 minutes per week in order to focus on reading and math.²⁷ A 2007 survey of 1,250 civics, social studies, and government teachers reported that 75 percent of respondents cited standardized tests as the reason current events were taught less frequently.²⁸ A five-year study completed by the University of Maryland in 2007 found "the pressure teachers were feeling to 'teach to the test'" since NCLB was passed has led to "declines in teaching higher-order thinking, in the amount of time spent on complex assignments, and in the actual amount of high cognitive content in the curriculum."²⁹

4. High-stakes testing is a detriment to differentiated instruction

A result of the one-size-fits-all standardized testing is that teachers are limited in not only what but also how they can teach. Each year, thousands of teachers are faced with a myriad of different learners in a single classroom that require a broad range of instructional methodologies and creative learning strategies in the classroom. However, in the current high-stakes testing climate, it is increasingly difficult for teachers to differentiate instruction or teach critical thinking skills to their students.

²⁴ The American Conservative, *Texas vs. No Child Left Behind* (January 23, 2013).

²⁵ FairTest's Testimony to the Rhode Island Legislature on Graduation Exams," www.FairTest.org, Feb 26, 2014, citing Koretz (2005).

²⁶ Jonathan Kozol, *The Shame of the Nation: The Restoration of Apartheid Schooling in America*, 2005

²⁷ Diane Ravitch, *The Death and Life of the Great American School System*, 2010.

²⁸ "Survey Finds Teaching to the Test Has Negative Impact on Use of News in Classrooms: Carnegie-Knight Task Force Urges More Emphasis on Civics Education," www.knightfoundation.org, Jan. 12, 2007.

²⁹ Bruce Jacobs, "No Child Left Behind's Emphasis on 'Teaching to the Test' Undermines Quality Teaching," *Endeavors*, Dec. 2007; Linda Valli, and Robert Croninger, "High Quality Teaching of Foundational Skills in Mathematics and Reading," drdc.uchicago.edu (accessed June 20, 2011).

Teachers often lack the authority to deviate from the standard curriculum regardless of student needs.³⁰ Teachers are generally expected to rigidly present a pre-programmed, often tightly scripted curriculum, covering a set of skills each day in order to prepare students for state assessments.³¹ Those who drown in this rigid, fast-paced system of one-size-fits-all education are the students with developmental delays or intellectual disabilities or those who are simply behind academically.

5. High-stakes testing causes significant student anxiety

Cases of students' experiencing significant test anxiety due to the high stakes associated with standardized tests, including the threat of not advancing to the next grade level, have been widely covered. The *Sacramento Bee* reported on March 14, 2002, that "test-related jitters, especially among young students, are so common that the Stanford-9 exam comes with instructions on what to do with a test booklet in case a student vomits on it."³² The known anxiety among students taking high-stakes standardized tests also results in an inaccurate measure of student performance.

6. High-stakes testing can cause a hostile school environment

With the frantic focus on student achievement measured by high-stakes tests, there is a fear that "some students" will bring down school performance and affect school ratings, a mindset that contributes to a stressful, often hostile school environment.³³ This type of environment produces behaviors that justify "zero tolerance" policies and sometimes lead to unnecessary suspensions, expulsions, and other disciplinary measures. Students caught in this web of frustration and detrimental behavior are more likely to drop out.³⁴

7. High-stakes testing puts pressure on teachers and students to cheat

Some believe the overemphasis on test scores in our education accountability system puts such extreme pressure on students, teachers, and administrators that cheating may result. Across the country, there have been reports of cheating on high-stakes tests, including some incidents here in Texas.

In 2014, one of Dallas's top-rated elementary schools based on students' performance on the state test, Umphrey Lee Elementary, was caught in a cheating scheme on Texas's 2012-13 test. Their results

³⁰ Neill, M., "A Child is Not a Test Score: Assessment as a Civil Rights Issue," *Root & Branch*, Vol. II, Issue 2, Fall 2009.

³¹ *Id.*

³² Susan Ohanian, "Collateral Vomitage," www.susanohanian.org, Mar. 14, 2002.

³³ "Racial Justice and Standardized Educational Testing," www.FairTest.org (downloaded May 2014), citing Advancement Project (2010), *Test, Punish, and Push Out: How "Zero Tolerance" and High Stakes Testing Funnel Youth Into the School-to-Prison Pipeline*.

³⁴ *Id.*

were found to be manufactured by teachers feeding students answers on most of the 2012-13 tests.³⁵ And in 2012, the El Paso ISD superintendent was accused of overseeing a six-year cheating scheme that included allegations of transferring low-performing students to charter schools, instructing students not to attend school on testing days, and reclassifying students in order to avoid tests mandated for sophomores.³⁶

Recommendations moving forward

Standardized testing, like many other things, can be useful in moderation. In the case of standardized testing, taking a periodic snapshot of the system as a whole can be useful as one of many tools policymakers use to gauge the health and performance of the overall system and some of its constituent parts. Because the benefit and validity of these tests are greatest at a system (state) level, they can and should be implemented differently from the current NCLB-driven testing regime. Specifically, the test should be low stakes, be administered less frequently, employ sampling, and be truly criterion-referenced.

Low-stakes tests are those that do not carry consequences for specific students, educators, or schools. The Texas test, for example, currently has direct punitive consequences for students and schools insofar as it is a grade promotion and graduation requirement for students and it is the overwhelming input for the state accountability system. Additionally, the Obama administration is pushing hard to get states, including Texas, to tie teacher evaluations and employment decisions to student test performance as a condition for receiving ESEA flexibility.

The current regime tests every student every year in multiple subjects. Valid, entirely adequate, system-level conclusions, such as those provided by NAEP, can be reached by testing a representative student sample of as little as five percent and alternating subject areas from year to year. As with NAEP, an appropriate sample will also allow for the disaggregated data that has been the one major success of NCLB.

Unlike a norm-referenced test that measures a student's performance against his or her peers, criterion-referenced tests measure students on whether they have mastered particular skills or standards. For example, a norm-referenced test would tell you that Sally can perform more of the skills associated

³⁵ Weiss, J. & Haag, M., "STAAR Scores, Rating Plunged at Top Elementary after Dallas ISD Determined Teachers Had Been Cheating," *Dallas Morning News*, Aug 14, 2014 (retrieved at <http://www.dallasnews.com/news/education/headlines/20140814-staar-scores-plummet-at-top-performing-disd-elementary-after-cheating-uncovered-among-teachers.ece>).

³⁶ Fernandez, M., "El Paso Schools Confront Scandal of Students Who 'Disappeared' at Test Time," *New York Times*, Oct 13, 2012 (retrieved at http://www.nytimes.com/2012/10/14/education/el-paso-rattled-by-scandal-of-disappeared-students.html?pagewanted=all&_r=0)

with a particular set of standards than John but less than Paul; however, the test would not tell you which or even what percentage of those standards Sally knows. A criterion-referenced test should tell you with some measure of validity that Sally can perform x percent of the skills associated with the standards being tested. Moreover, the test can tell you with a great degree of validity that all of the students who took the test can perform x percent of the skills tested. As stated above, when the population of test takers is a statistically valid sample of the total student population, conclusions drawn from the sample can be made about the student population as a whole and about substantial subpopulations.

Recommendation 1: Allow individual states to create their own test or use the NAEP, a state-level oversample of PISA, or another valid multisubject international benchmark, and report disaggregated state-level information derived from testing either the entire student population or a scientifically valid sample of the student population.

Example: Non-test based assessment

One of the main questions asked by the paper this testimony was drawn from was, “What are the realistic consequences of Texas policymakers choosing to simply walk away from testing under NCLB?” One option that we explored was to continue assessing every child every year in grades 3–8 and at least once in grades 9–11 but to do so without using a system based on standardized testing. The following is an example of one of many potential ways to assess students using a standardized process that does not involve a traditional standardized test.

Example: Peer-to-peer assessment system

The following is an example of a potential standardized system of assessment not based on traditional standardized testing; other examples exist. Under the system, Teacher A would input all or a sample of her students’ work (worksheets, tests, pop quizzes, writing assignments, or other miscellaneous projects) into a statewide database or portfolio system. The system would randomly select 5–15 percent of the assignments given to Teacher A’s class; request that Teacher A provide a rubric for how the assignment was initially graded; and randomly assign that work and the rubric to Teacher B to be re-graded. Optimally, Teacher B is on a different campus from Teacher A and the group from which Teacher B is selected is composed of master’s-level teachers. Teacher B will re-grade the work without knowledge of the grade assigned by Teacher A but using the same rubric. After the re-grade is completed, Teacher B will enter the second grade into the assessment data system. Assuming Teachers A and B give a student the same or near-equivalent grade, the grade on that student’s work will be considered validated. If the grade Teacher B assigns varies significantly from the grade given by Teacher A, the system will assign that student’s work to Teacher C to perform an additional re-grade. After Teacher C resubmits the grade

into the assessment data system, the system will compare the grade given by Teacher C to the grades given by Teachers A and B and validate the two grades that most closely match. At the end of the year, all the students' validated scores will be averaged together for each relevant course, and that average score will be the assessment score used to satisfy NCLB's assessment requirement.

In addition to providing a deeper, more meaningful depiction of student performance, this system also has the benefit of improving the teaching profession through continuous feedback on whether each teacher is accurately assessing student performance as validated through continuous peer review. Again, this example is but one of many possible ways to develop a standard system of student assessment that does not rely on traditional standardized testing.

Whether or not the US Department of Education would accept such a system as fulfilling the current mandates in NCLB is unknown. However, through its reauthorization of NCLB, the legislature could make it clear that such an innovation would satisfy the requirements under the law.

Recommendation 2: Encourage states to pursue educational innovation and more authentic forms of student assessment by expressly allowing them to implement a standard system of student assessment that is not based on standardized testing.