



601 PENNSYLVANIA AVE, N.W.  
 SOUTH BUILDING, SUITE 900  
 WASHINGTON D.C., 20004  
 PH: 866-240-7909

**BOARD OF DIRECTORS**

- \*Andre Allen*, General Dynamics Information Technology
- Douglas Becker, Ph.D.*, HMH Riverside
- Nikki Eatchel*, Scantron
- Henrik Skovdahl Hansen, Ph.D.*, Dansk Psykologisk Forlag
- Christine Niero, Ph.D.*, Professional Testing
- John Oswald*, ETS
- Dan Rinn*, National Technology Transfer
- Rachel Schoenig, J.D.*, Cornerstone Strategies, LLC.
- Cicek Svensson*, Cicek Svensson Consulting
- Alex Tong*, ATA
- Timothy Vansickle, Ph.D.*, Questar
- John Weiner*, PSI Services LLC
- Bill West*, Pearson VUE
- \*Chair*

**Chief Executive Officer:** *William G. Harris, Ph.D.*

**General Counsel:** *David Arnold, Ph.D., J.D.*, Wonderlic, Inc.

**Secretary:** *Gerald Borofsky, Ph.D.*, Bay State Psychological Associates

**Treasurer:** *Walter Wang*, Collegepre

COMMENTS OF THE  
 ASSOCIATION OF TEST PUBLISHERS

Response to the Notice of Proposed Rulemaking  
 On Implementing Title I of the Elementary and Secondary Education Act—  
 Improving the Academic Achievement of the Disadvantaged—  
 Academic Assessments

[Docket ID ED-2016-OESE-0053]

July 11, 2016

---

The Association of Test Publishers (“ATP”) provides these comments in response to the Notice of Proposed Rulemaking issued by the United States Department of Education (“Department”), dated July 11, 2016 (81 *Fed. Reg.* 44928), seeking comments on the proposed consensus Title I assessment regulations developed by the Negotiated Rulemaking Committee to implement the Every Student Succeeds Act (“ESSA”), amending Title I of the Elementary and Secondary Education Act. Those proposed regulations amend the current regulations in 34 CFR Part 200. The ATP’s comments and recommendations are submitted timely by the due date of September 9, 2016.

The ATP is the international trade association representing approximately 175 publishers, developers, and vendors of assessments and assessment services (both non-profit and for profit) used in a variety of settings, including virtually every educational purpose for which the Department of Education is responsible. The ATP serves as the “Intelligent Voice for Testing,” providing input to the United States Congress, state legislatures, and federal and state agencies and courts, in their efforts to examine and resolve issues surrounding testing and the use of test data.

These comments are submitted on behalf of the ATP's educational assessment members, including many of the entities that have historically provided testing products and services to the states under the Improving America's Schools Act ("IASA") and No Child Left Behind Act ("NCLB") and related laws and regulations administered by the Department – and who will be called upon to assist in implementing the ESSA. Moreover, ATP members are called upon to provide evidence used in the peer review process, and to serve as a resource to educators at every level on a variety of assessment-related issues. Accordingly, the ATP feels it is critical that the transition from NCLB to ESSA be smooth and free of debate as states and districts continue to move forward with their assessments, assessment systems, and/or assessment programs.

The members of the ATP have decades of experience in developing and implementing complex assessment systems in all 50 states, plus the District of Columbia, US Territories and DOD schools, and in many of the nation's 15,000 school districts. These testing companies work closely with their SEA clients and the consortia to ensure that statewide testing programs are implemented and operated in accordance with all federal and state regulations. They also work equally closely with LEAs in pursuing and implementing effective local assessment strategies. The U.S. testing industry comprises educators, researchers, psychometricians, and technologists with extensive experience in developing and administering technically sound assessments that are used for many different purposes.

Significantly, the ATP has collaborated with the Council of Chief State School Officers since 2007 in developing and maintaining best practices and related guidance for states and districts in the administration of statewide assessment programs. The original version of the *Operational Best Practices for Statewide Large-Scale Assessment Programs* was published in June of 2010; an updated version adding a new focus on technology-based assessments and accessibility, including those used for summative, interim, and formative purposes, was published in August 2013 (hereinafter referred to as "*Operational Best Practices*"). This guidance is now considered as the seminal documentation and training guide for states and test publishers; it was cited with approval by the U.S. Government Accountability Office in its report entitled, *K-12 Education: States' Test Security Policies and Procedures Varied*, (GAO-13-495R, May 16, 2013), as the benchmark in evaluating state security plans for addressing testing irregularities, security breaches, and cheating by test takers and school officials. Although the *Operational Best Practices* are proprietary, this resource reflects the collective expertise of the testing industry and is appropriate for the Department to consider in resolving issues about the final Title I A assessment regulations.

As the *Operational Best Practices* demonstrate, the testing industry takes its relationship with its customers (as well as with student/parent stakeholders) seriously, seeking to provide professional assistance in every aspect of the development of assessments to demonstrate they are fit for purpose, as well as how they are delivered, administered, scored, and how scores are interpreted and reported. Test publishers also meet the professional technical standards embodied in the *Joint Standards for Psychological and Educational Testing (2014)*, developed and issued by the American Psychological Association, the American Educational Research Association, and the National Council for Educational Measurement (hereinafter referred to as the "*Joint Standards*"). Based on these standards and best practices, states usually engage in robust RFP processes to select the very best assessments (typically from various sources) they feel can meet the requirements that they themselves have established to serve the needs of their programs and their students. Thus, publishers have developed tests in direct response to state RFPs that set out detailed descriptions of what each state is seeking and what it wants to include

in its assessment systems. Unfortunately, sometimes funding constraints may limit the scope of assessment systems and the capability to implement available innovations.

Many education experts have noted that while the ESSA makes significant changes to federal requirements, it maintains the requirement to use “high quality “valid and reliable statewide assessments in at least reading/language arts and mathematics in each of grades 3-8 and once in high school; and in science in each of three grade spans (3-5, 6-9, and 10-12). The ESSA also adds significant flexibility, including the ability to use locally-selected high school assessments and innovative assessment systems, and allowing states to measure student achievement and growth “through a single summative assessment” or “through multiple, statewide interim assessments during the course of the academic year that result in a single summative score.” A further innovation in the law provides a pilot for competency-based and/or computer adaptive assessments. While these options are not without psychometric challenges, the testing industry looks forward to providing valuable assistance to states and districts so that new assessment systems based on ESSA requirements are fair, valid and reliable, and produce the best possible data for the evaluation of student and school performance.

The ATP applauds the Rulemaking Committee for advancing state control and flexibility, while adhering to the requirement that all assessments remain technically sound – and thus avoiding any major conflicts with statutory language. The US testing industry looks forward to providing assistance to the Department and in working with states and districts so that new assessment systems based on ESSA requirements are fair, valid and reliable, and produce the best possible data – which in turn will continue to provide appropriate test data for use in evaluating student performance. Accordingly, the ATP appreciates the opportunity to provide the following comments and recommendations to clarify the proposed regulations.

## **COMMENTS**

### **1. Observations About Assessment Innovation**

Although the term “innovative” is not used in the proposed consensus regulations, that concept continues to be reflected in the urgency expressed among stakeholders that ESSA assessments must be innovative. As the Department noted in its testing guidance letter to Chief State School Officers, dated February 2, 2016, the ESSA “takes additional steps to support smart, effective assessments ... and to strive for continued improvement and innovation in assessments.”

ATP members pride themselves on the ability to innovate, whether that has been in the context of individual state/local programs or as part of the various assessment consortia. Indeed, the industry has responded quickly and energetically to assist states to implement a wide variety of education reforms and policies. In so doing, the ATP and its members have been in the forefront of efforts to utilize technology to advance the goals of education reform. For example, the industry has pioneered the development and use of:

- Performance-based and portfolio assessments, in addition to constructed response and essays;
- Technology-enhanced items (e.g., drag and drop, hot spot, gamification);

- Formative and interim assessments;
- Technology-based student assessments administered online, as well as for using paper and pencil with technology-based scoring and assembly;
- Technology-based adaptive assessments based on an item bank, that enable each student to be assessed accurately and quickly using items best tailored to measure that student’s level of knowledge and achievement;
- Vertical scaling and growth measures, which in fact predates the emphasis on alignment of standards in the Improving America’s Schools Act and later in the No Child Left Behind Act, as well as the current focus on growth indicators;
- Tests that provide BOTH normative and criterion-referenced interpretations of student performance;
- International benchmarking;
- Tests of college and career readiness;
- Assessments for English learners and students with disabilities that adhere to universal design principles; and
- Extensive and sophisticated data and reporting systems that allow districts, principals, teachers, and parents to monitor and predict student performance early enough to target interventions and resources to meet individual student needs so they can meet college and career ready standards.

Another area of special innovation that the testing industry has undertaken is the development of a comprehensive set of best practices for operating statewide testing programs. These best practices, which were developed jointly by the ATP and the Council of Chief State School Officers (CCSSO), complement the *Joint Standards*, which address psychometric properties of tests and the technical aspects of measurement and assessment. These best practices cover every element of statewide programs, from the RFP stage to program management, to item banking, to administration and test security, to scoring and reporting of test data, and include best practices for technology-based assessments (e.g., computer-based, computer-adaptive, linear-on-the-fly) and the use of accessibility principles to assess special populations.<sup>1</sup>

The ATP recommends that the Department formally acknowledge innovation as a goal of the proposed regulations.

---

<sup>1</sup> In their *Operational Best Practices*, the ATP and CCSSO have defined the term “technology-based assessments” to clarify that “online” is not a proper characterization. Today’s technology has led to the use of many student assessments that are administered using a computer (or other device) that are not administered online with direct access to the Internet, but rather where the student accesses digitally-formatted assessments directly, remotely through a server that caches the assessment, or through some storage medium (e.g., CD, USB flash drive).

## **2. The Requirement for Technical Quality of Assessments**

The proposed regulations stress the need for maintaining technical quality of assessments by requiring that a state implement a “system of high quality assessments” and requiring that assessments “are of adequate technical quality.” See §§200.2(a)(1) and (b)(4)(1). They also require that Title I assessments must be “consistent with relevant, nationally recognized professional and technical testing standards. See §200.4(ii).

The US testing industry has supported the use of high quality assessments for decades, principally by relying on the *Joint Standards*, which was first published in 1966 (the most recent edition of which was published in 2014). The *Joint Standards* represents the gold standard in professional, technical guidance on assessment in the United States and in many other countries – a fact that has been recognized by federal, state, and district education policymakers. Indeed, Congress has referred to the *Joint Standards* as the definitive benchmark for assessments many times (e.g., Goals 2000, the Improving Americas Schools Act, No Child Left Behind, and several versions of IDEA reauthorizations).<sup>2</sup> Congress has continued this commitment to the *Joint Standards* in its requirement that Title I assessments must be valid, reliable and fair. See Section 1111(b).

Therefore, the ATP can only assume that the requirement that Title I assessments must be “consistent with relevant, nationally recognized professional and technical testing standards” must refer to the *Joint Standards*. Other provisions in the proposed regulations also track the *Joint Standards* -- and provide clear guidance to states and districts as to exactly what professional technical standards its assessments must meet. Examples include: (a) §200.4(i) would require that assessments must be “valid, reliable and fair for the purposes for which the assessments are used”; (b) §200.7(i) would require that measures of student academic achievement “include valid and reliable measures of student academic growth”; and (c) §200.3(b)(1)(v) would require that nationally-recognized high school academic assessments must “produce valid and reliable data on student academic achievement.” See also §200.4(b)(1)(ii), where parallel requirements are set forth where a state law exception exists.

The ATP notes, however, that there are some inconsistencies in the usage of the appropriate *Joint Standards* terminology, where non-*Joint Standards* terminology is mixed in, such as “accessibility” or “transparent.” See §200.2(b)(2)(i) and §200.2(b)(10)(ii). Those other terms, while having some application in the regulations, should NOT be used as though they are part of the traditional *Joint Standards*. As currently written, the proposed regulations are likely to cause confusion. For this reason, to provide clarity, the ATP recommends that the proposed regulations should be revised so that they are limited to the phrase “valid, reliable, and fair.”

## **3. The Requirement for Universal Design.**

The ATP believes a similar potential for confusion exists in the proposed language in §200.2(b)(2)(i), which requires that Title I assessments be “developed, to the extent practicable, using universal design for learning” as defined in subpart (ii). Although the ATP recognizes that

---

<sup>2</sup> It bears noting that the *Joint Standards* also are the foundation for the Department’s Peer Reviewer Guide. The ATP will address peer review below in comment number 6.

the term “universal design for learning” as articulated by the National Center on Universal Design, is consistent with the ESSA language, and while the proposed regulation language is well intended, this concept is less directly related to the assessment science addressed in these regulations than other, more appropriate, research-based sources. Therefore, the ATP recommends that the universal design concept – and thus the requirements -- should be specifically directed at assessments. In order to achieve this result, the ATP recommends that the regulation include a focus on test design frameworks that are based on universal design principles.

The ATP/CCSSO *Operational Best Practices* directly addresses the process for using universal design in the context of a large-scale assessment. Of critical importance, this guidance stresses that universal design principles are to be applied at the beginning of and throughout the entire test development process, not merely factored in at the end on after test items have been designed. As stated (Introduction to Chapter 2, page 37), a state should use well-defined test specifications, including universal design, to develop and review assessment items.

In fact, the *Operational Best Practices* provides a number of specific details as to how this approach should be carried out. See Chapter 19 (“Assessment of Special Populations/Accessibility for All Learners”).<sup>3</sup> Specifically, Section 19.1 states:

A process should be established, implemented, and monitored/evaluated to ensure that test items, forms, and pools are developed in accordance with the principles of universal design and sound testing practice for both paper-based and technology-based assessments.”

The section then goes on to identify and discuss seven areas where universal design principles in assessment come into play, including the need for a test specification addressing: (1) organization and sequencing of items; (2) presentation of items, including rules for alternate text (e.g., tagging, audio presentations); (3) positioning of items and passages and/or other stimuli; (4) font and point size; (5) margins and blank space; (6) navigation within and between items; (7) ways to indicate answers, such as input modes, response modes, and technology-based response modes (e.g., mouse, keyboard, touch screen, assistive device); (8) timing requirements, if applicable; (9) clarity of graphics and/or item stimuli; (10) use of color and shading (e.g., highlighting); and (11) rules for emphasizing words or phrases (e.g., bolding, capitalizing, underlining and italicizing, audio amplification and/or speed). The Department may find it useful to consider this guidance in modifying the proposed regulations.

The ATP’s focus on the appropriateness and benefits of universal design in assessment actually pre-dates the *Operational Best Practices*. In 2004, the ATP published a paper noting that universally designed assessments are developed to allow participation of the widest possible range of students, in a way that results in valid inferences about performance on grade-level standards for all students who participate in the assessment. That paper explored the development of universal design and considered its application to large-scale assessments. See

---

<sup>3</sup> See Association of Test Publishers and Council of Chief State School Officers, *Operational Best Practices for Statewide Large-Scale Assessment Programs*, 165-68 (2013).

Thompson, S, Thurlow, M, and Malouf, D, “Creating Better Tests for Everyone Through Universally Designed Assessments,” *Journal of Applied Testing Technology*, 2004 Vol. 1.<sup>4</sup>

For these reasons, the ATP recommends that the Department clarify §200.2(b)(2)(i) to expressly refer to “universal design of assessments” and provide appropriate guidance as to its meaning.

#### **4. The Requirements for Evidence.**

Under the proposed regulations, states are required to submit “evidence and/or use “evidence-based” practices in various situations, each of which impacts how a state drafts and implements its state plan and effective strategies which have either demonstrated effectiveness or show significant potential. For example, §200.2(b)5 requires that ESSA assessments must be supported by adequate evidence of technical quality, while each state must submit evidence that its assessments, including any locally-selected nationally-recognized high school academic assessments, meet all of the requirements of the ESSA for peer review under Section 1111(a)(4). *See* §200.2(d) and §200.3(b)(ii).

The proposed regulations do not offer any guidance as to what constitutes evidence in these situations. The ATP recommends that the Department, at a minimum, clarify that the technical manual developed for any large-scale assessment should be considered as appropriate evidence in these settings.

The technical manual for any large-scale assessment provides information about the technical quality of assessments, as well as evidence related to universal design. Such manuals typically include information on the intended purpose and valid use of the assessment, how the assessment was developed, administered, scored, interpreted, and reported, and additional details about any technical studies done on the completed assessment.

#### **5. The Requirement for Multiple Measures.**

The proposed regulations (§200.2(7)), would require that state assessments “involve multiple up-to-date measures of student academic achievement.” Fundamentally, the US testing industry has endorsed the concept of multiple measures since the early days of education reform. In its testimony before the House Committee on Education and Workforce in 1988 on Goals 2000, testing representatives struck the chord of harmonizing strong technical quality with the need for multiple measures of what students can do and for ensuring that teachers and school building leaders obtain useful information from those assessments in order to inform teaching and learning. Moreover, in advocating for multiple measures, the testing industry has historically been committed to the concept that assessment systems must be built based upon an identified purpose or purposes of each assessment in the system. *See* §200.2(b)(2)(i).

The ATP has repeatedly advocated on the value of multiple measures and how the use of multiple measures can be taught. The testing industry aggressively worked with the Department’s Office for Civil Rights in 2000 as it developed guidance for high-stakes testing. In a related vein, the ATP has urged Congress to include professional training for teachers and

---

<sup>4</sup> At the time of publication of this article, one of the authors, David Malouf, was employed by the U.S. Department of Education.

principals on the use of assessments and assessment data as an allowable use of funds under the Higher Education Act (“assessment literacy”) – to ensure that those leaders in the school and classroom receive more than in-service training in the use of assessments and assessment data.

The ATP also supports the proposed regulations requirement that Title I assessments be designed for the purposes for which they have been identified. Indeed, the *Operational Best Practices* clearly articulate that the state/district “must clearly understand and articulate the purposes of the new assessment design and the intended assessment data uses.” See Pre-Chapter §III(c) (page 18).

## **6. The Requirements for Peer Review.**

The proposed regulations (§200.2(d) and §200.3(b)(2)(ii)) would continue the long-standing requirement that a state submit evidence about its assessments/assessment system for peer review. The peer review process has been in place for more than 20 years, since implementation of the Improving America’s Schools Act of 1993, which established the core requirements for statewide assessments under the Elementary and Secondary Education Act. Since its inception, peer review has been the means to evaluate state assessment systems. As the Department recognized in its most recent changes to the procedures, peer review applies equally to each state assessment program, regardless of what assessments are being reviewed or who developed those assessments. Consequently, as the ATP has previously commented, the peer review process must be transparent, fair, and free from all bias and subjective decisions, adhering to objective and technically-sound principles.

Accordingly, as the Department considers these new regulations, the ATP reiterates that the Department must ensure that a process exists by which to evaluate on a consistent and objective basis the technical quality of state assessments to ensure their reliability and validity, regardless of whether a state is a member of an assessment consortium. In that sense, the long-standing requirements for states to ensure that their assessments are valid and reliable under the *Joint Standards* remain in full force and effect. We reiterate that there should be no reason to prefer a consortium assessment over any other statewide assessments or to prefer an assessment merely because some assessments contain particular item types (e.g., technology enhanced items). In reality many state assessment systems already use multiple item types, including technically enhanced items, performance-based tasks, and well-developed multiple-choice items that are capable of assessing depth of knowledge.

On the issue of the selection of peer reviewers, the ATP reiterates that the process should ensure each panelist is vetted to confirm that he/she can conduct an assessment evaluation without any appearance of a conflict of interest and that each individual possesses the requisite expertise and qualifications. The ATP also recommends again that training should be available to all panelists to ensure that these objectives are met. Once a panel is selected to review a state submission, the same panel should be used throughout the review process, instead of allowing different panel(s) during state resubmission following an initial rejection, leading to a “shifting burden” for a state to meet as new reviewers’ opinions come into play. Finally, the peer review process must provide appropriate and adequate due process for a state to challenge the decision reached by the panel. Such due process should include the right to petition for reconsideration of a peer review decision, as well as setting forth a process by which a state is entitled to seek judicial review of a non-approval decision. This composition, vetting, and training would ensure that evidence submitted by a state is reviewed consistently between panels, by truly objective reviewers who



possess all of the relevant expertise and skill sets needed to make a full, accurate, and impartial evaluation of each assessment system.

## **7. The Requirements for Alternate Assessments and Other Accessibility Issues.**

The proposed regulations limit the use of alternate assessments to 1% of students (i.e., students with severe cognitive disabilities), though the state may not prohibit an LEA from exceeding that percentage. Instead, a state may require that a LEA submit information justifying its need to exceed 1% alternate assessments, making that information publicly available (so long as it does not reveal personally identifiable student information), and may provide “appropriate oversight.” Additional regulatory language requires that states with alternate assessments establish guidelines for IEP teams to use in determining which assessment should be used, and that students should not be identified for such assessments on the basis of a particular disability, previous low achievement, or status as an English Learner.

Moreover, the proposed regulations place a heavy reliance on what are “appropriate accommodations” for the statewide assessment based upon the student’s IEP, or recommendations from the student’s placement team or the LEA team. All of these approaches tend to individualize the process of accommodation.

To ensure that the purpose and operation of these related proposed regulations is clear, the ATP recommends that the Department clarify that all assessments, including alternate assessments, should be standardized assessments that adhere to the *Joint Standards* and are of high quality, fair, and reliable, and produce valid results and interpretations. In the past, too many alternate assessments were individualized, often with the mistaken impression that they would better meet the requirements of IDEA. Any suggestion that states should use non-standard, non-objective measures for the alternate assessment (or for that matter the general assessment) is inappropriate. All assessments must support decisions about students with disabilities that are appropriate and valid. Equally important, comparability among a state’s alternate assessment and its general assessment is central for a state’s accountability decisions. The ATP strongly believes that standardized alternate assessments with demonstrated comparability are being developed today that will best serve students with disabilities. Accordingly, the ATP recommends that the proposed regulations require that test developers adhere to the *Joint Standards*, universal design principles, and other applicable assessment best practices, in constructing alternate assessments, so that special populations, including students with the most significant cognitive disabilities and English Learners, are measured by high quality, fair, reliable, and valid assessments.

In addition, the ATP recommends that where “appropriate accommodations” are allowed, that the Department should clarify the regulations to identify that a state should develop with its assessment vendor(s) a list of accommodations that have been determined to be capable of being used with the statewide assessment without jeopardizing the validity and reliability of any score interpretations of students with disabilities. Moreover, a current list should be shared with parents of students and the IEP and LEA placement teams so everyone concerned is acquainted with this information well in advance of any testing period.

## **Conclusion**

The ATP appreciates the opportunity to provide comments on the proposed Title I assessment regulations. We hope the Department will find these comments and recommendations useful to help improve our educational system and ensure that state and local assessments/assessment systems function smoothly and effectively to meet the needs of all stakeholders.

The ATP has at its disposal a wide array of psychometric content, alignment, and design resources and we stand ready to respond further on these issues if the Department has any questions or requests for further information.

Sincerely,

A handwritten signature in cursive script, appearing to read "William G. Harris".

William G Harris, Ph.D.  
CEO