Maximizing the Effectiveness of Online Accountability Assessments for Students with Disabilities

Mary Axelson

Online assessment promises a faster and more useful return of data about student performance in states’ accountability assessments. The prospect of gaining quick access to such information is alluring, and many states are creating new assessments for an online environment in order to obtain it. This period of retooling presents a significant additional opportunity— that of improving the assessment of students with disabilities.

By employing universal design principles—i.e., using methods that take differing abilities into consideration from the start of a project—states can maximize the effectiveness of their new assessments for measuring the knowledge and skills of students with disabilities, while also improving the speed and usefulness of their systems generally.

Adapting universal design practices can guide assessment developers as they work to capture the considerable potential of online environments to give educators a better picture of student learning. Universal design of assessment is specifically intended to benefit those students who have physical disabilities or learning disabilities that impede their interpretation of assessment items or their responses to them. Secondary benefits for all students, however, are likely to be the final result. This brief outlines the potential of acting during this historical moment to significantly improve accountability assessments through universal design of their online versions.

Success with universal design will require collaboration among individuals with expertise in several distinct areas of specialized knowledge— including universal design, special education, online technology systems and applications, assistive technology, assessment, and content standards— all working within an environment of evolving legislation. This document provides an overview of the various facets of this congruence of specialties, discusses the potential of their interplay, and encourages a joint effort as states create online assessments.

Why Seek Change?

Using state accountability assessments to measure the academic progress of students with disabilities is a relatively recent development. Until 1997, students with disabilities were routinely excluded from large-scale assessments, and reporting of scores for disabled students who did participate varied widely.1
Advocates for students with disabilities have long argued that if scores for students with disabilities are not reported separately from those of other students, individual students are underserved, the performance ratings of schools are less meaningful, and schools lose an incentive to focus attention on the academic performance of this population.\(^2\)

Both the 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA) and the accountability requirements of the No Child Left Behind Act (NCLB) now require states to assess students with disabilities with the statewide test and to report those scores in both aggregate and disaggregated forms.\(^3\) Additionally, failure to report scores violates section 504 of the Rehabilitation Act of 1973 and Title II of the Americans with Disabilities Act of 1990. Title I allows up to 1 percent of students with the most severe cognitive disabilities to take an alternate assessment, and in the spring of 2005, the U.S. Department of Education announced its intention to grant states additional flexibility with another group (2 percent of students) “in need of modified standards and assessments.”\(^4\)

Current mandates for testing and reporting the performance of students with disabilities put states under more pressure than ever to make sure assessments provide accurate information about what such students know and can do. Yet a number of disabilities can prevent an assessment from accurately evaluating a student’s knowledge and skills. If a student has difficulty interpreting text, for example, print-based assessments may evaluate reading instead of mathematics, science, or other intended targets of the assessment.

### An Introduction to Issues of Online Assessment

The online, multimedia delivery of assessment can offer greater flexibility in interaction than print-based assessments. That flexibility can enable assessments to be more closely matched with the diverse ways in which students receive, process, and respond to information.

Software and hardware tools can, for example, allow students to select an answer by gazing at the screen, to change the appearance of text, or to answer an open-response item by simply speaking—and that audio can be analyzed by computer. Minute neural or muscle movements can initiate computer commands. Prompts can help students remember to use learned strategies for organizing their work or kicking their memory into gear. Technology-based assessment holds stunning potential for administering tests, engaging students, diagnosing learning styles and disabilities, and immediately capturing and handing back data on the effectiveness of school programs. New item types, such as editing a document, making a concept map, manipulating a simulation, or drawing a river on a map, are also emerging.

Yet, as promising as online testing is, the experiences of some states have shown that making even the most basic conversion to online testing, i.e., simply replicating a paper-based test for digital delivery, is an enormous task. Large text, for example, is not an improvement if students must scroll sideways to read it. Furthermore, reformating existing items for presentation on a screen and opening them to a greater number of assistive technologies creates new potential for affecting the reliability and validity of an assessment. For example, a paper-based test allows several items to be seen at once, and it is not known whether presenting a single item affects context in a way that could influence performance.

Recently, the Appalachia Educational Laboratory at Edvantia and the Council of Chief State School Officers convened interviews and panel discussions on creating online assessments of technology skills. These conversations revealed that assessment developers are finding that creating online assessments requires a different design process than creating paper-based tests. In their experience, the traditional tag-team approach to test development needs to be replaced by greater collaboration.\(^5\) This need for a new design process presents yet another opportunity to adopt a universal design approach that involves a variety of specialists.

### An Introduction to Universal Design

Considering the needs of all potential users from the inception of a project is central to the philosophy of universal design. Retrofitting for disabilities is not universal design. The philosophy gained clout in architecture with passage of the Americans with Disabilities Act. Universal design’s relationship to technology was pushed forward by the Assistive Technology Act (ATA) of 1998, which defines it as

a concept or philosophy for designing and delivering products and services that are usable by people with the widest possible range of functional capabilities, which include products and services that are directly accessible (without requiring assistive technologies) and products and services that are made usable with assistive technologies.\(^6\)

The 2004 reauthorization of the Individuals with Disabilities Education Act references the ATA definition above and specifically calls for universal design in regard to access to the general curriculum, to assistive technologies, and to assessment. The language pertaining to assessment states, “The State educational agency (or, in the case of a districtwide assessment, the local educational agency) shall, to the extent feasible, use universal design
principles in developing and administering any assessments under this paragraph.”

Universal design of assessment is an emerging extension of a growing body of work in universal design for learning (UDL). UDL began with guidelines for accessible media, such as recommendations for font size and color for students with visual impairments and interoperability with alternate keyboards for students with motor impairments. UDL now strives to apply findings from recent brain research. For example, the Center for Applied Special Technology (CAST) recommends that authors of standards, curriculum, and assessment consistently ask which of three neural networks (recognition, strategic, or affective) is to be addressed. (See the section of this brief headed “Universal Design and Accommodations” for details.)

Making assessment, as opposed to instruction, accessible to all students presents an entirely new set of challenges, especially in finding a balance in the use of accommodations and the design of assessment items—issues discussed next.

**Universal Design and Accommodations**

Universal design aims to embed or provide access to the widest possible range of accommodations. Many test items might require abilities other than knowledge of the construct being tested. A student may, for example, thoroughly understand the timeline of events in a reading comprehension passage, but answer incorrectly because she did not have the stamina in controlling her motor functions to fill in the bubbles to complete a series of questions. In such a case, allowing her to push buttons instead could provide a more accurate picture of her academic abilities.

A wide range of state policies, well documented by the National Center for Educational Outcomes (NCEO) and WETA public television’s LD Online Web site, addresses the use of accommodations to make assessments more accessible to students with disabilities. Examples include allowing extra time, providing an isolated environment, reading questions aloud, enlarging the print, making items available in Braille, providing prompts to keep students focused and organized, and allowing verbal responses. Use of such accommodations in instruction and assessment varies widely, as does school-level knowledge of the accommodations that are possible. Current practice gives the individualized education plan (IEP) team responsibility for allowing accommodations for an assessment, and states often provide lists of “allowable” accommodations from which the IEP team can choose.

Although liberal use of accommodations for assessment may simplify the test development process, such use is of great concern to psychometricians. These assessment experts safeguard the reliability, validity, and fairness of assessments, and accommodations for students with disabilities have an unknown impact on such measurements. It often is not known when an accommodation becomes an unfair advantage rather than a way to even the playing field.

In higher education, the right to a level playing field for assessment has been clarified with several legal decisions brought by Disability Rights Advocates (DRA, Oakland, California), a nonprofit legal center created to protect the civil and human rights of people with disabilities. K-12 decisions are not as common, but one K-12 lawsuit by DRA against the State of Oregon explored the impact of the state’s high-stakes assessment on students with disabilities on the Oregon Statewide Assessment System (OSAS). That case produced a widely referenced Blue Ribbon Panel Report, released in 2000, with recommendations on processes and policies for including students with disabilities on the state exam. Most significantly, this report offered guidance on how to set policy for the use of accommodations by students with disabilities (all of whom, by law, have an IEP) during this period when we lack clear research on how accommodations influence reliability and validity.

Accommodations should be allowable, valid, and scored if they are consistent with instructional and classroom accommodations included within a student’s IEP—unless and until research invalidates the construct and purpose of the OSAS (Oregon State Assessment System).

In a report on issues and evidence in assessing students with disabilities, Barton and Koretz observe that accommodations have been shown to significantly increase...
participation in large-scale assessments, even if their influence on the reliability and validity of test scores remains an uncomfortable unknown. Indeed, these authors point out that the very definition of the word has recently shifted. Accommodation once referred to changes in presentation, response, or setting, while modification referred to alterations of the actual assessment. The current edition of the Standards for Educational and Psychological Testing, they write, categorizes both as accommodation.\(^{12}\)

**Universal Design and Assessment Items**

While the use of accommodations will continue to be one aspect of universally designed assessment, another, perhaps equally powerful, aspect is the further refinement of assessment items.

Advances in brain research hold great promise for helping assessment authors create more precise items. Much of CAST’s work in UDL, as mentioned earlier, is based on the finding that the brain makes use of three distinct neural networks: recognition, strategic, and affective. CAST recommends that authors of assessment items ask themselves which network each item addresses. As explained in the book *Teaching Every Student in the Digital Age*,

- Learning is distributed across the interconnected networks: the recognition networks are specialized to receive and analyze information (the “what” of learning); the strategic networks are specialized to plan and execute actions (the “how” of learning); and the

---

**Existing Design Guidelines for Instructional Materials, Assessment, and Online Accessibility**

Universal design of instruction and instructional materials is a new movement; universal design of assessment is even less developed. Given that an established tenet of testing is that students be assessed with the same tools with which they learn, IDEA’s call for universal design in curriculum, access to assistive technologies, and assessment makes sense. Even so, technical standards and design guidelines lag behind the legislation.

Most current guidance relates to physical disabilities. For example, a number of excellent guidelines for designing computing environments for people with physical disabilities exist; these specify fonts, colors, navigation techniques, and other details. However, the vast majority of the nearly 6.5 million\(^{21}\) students served under IDEA do not have physical disabilities. Students with specific learning disabilities (a categorization used for IDEA reporting) constitute half of all students with disabilities in public schools.\(^{22}\) Students with speech or language impairments (18.9%), mental retardation (10.6%), and emotional disturbance (8.2%) are the next largest categories. According to the Office of Special Education Programs, “Together, these four categories represent 87.7% of all students ages 6-21 served under IDEA.”\(^{23}\) Design principles for these students are just emerging from a number of sources and are briefly described below.

**Technical standards for instructional materials.** Technical standards now in development hold potential to advance universal design of assessments. CAST has two five-year cooperative agreements from the Department of Education’s Office of Special Education Programs to establish two national centers to further develop and implement the National Instructional Materials Accessibility Standard (NIMAS). NIMAS guides the production and electronic distribution of digital versions of textbooks and other instructional materials so they can be more easily converted to accessible formats, including Braille and text-to-speech.\(^{24}\)

Technical standards specific to learning materials are beginning to address assessment. The IMS Global Learning Consortium develops and promotes the adoption of open technical specifications for interoperable learning technology (see www.imsglobal.org). IMS specifications include the IMS Question and Test Interoperability specification,\(^{25}\) which allows for such needs as an interoperable bank of test items. Informing the keepers of these standards about assessment needs for students with disabilities could be of great value. The Southern Regional Education Board (SREB), for example, points educators to IMS while adding recommendations of its own. One SREB recommendation specific to assessment is to organize test/question item banks by the learning outcomes they are intended to assess.\(^{26}\) Organizations that formulate standards have processes in place to consider such suggestions and act on them. The SREB recommendation, for example, may be accomplished with a metatag for learning outcomes on each test item.

**Guidance on assessment.** Key personnel of CAST, in a study on the effect of an accommodation on test results, state, “Testing accommodations such as the read-aloud have led to improvement, but research findings suggest the need for a more flexible, individualized approach to accommodations.”\(^{27}\) One route to flexibility employs multiple measures of abilities and multiple ways for students to work with assessment materials. Acknowledging the three neural networks
affective networks are specialized to evaluate and set priorities (the “why” of learning).

- Learners cannot be reduced to simple categories such as “disabled” or “bright.” They differ within and across all three brain networks, showing shades of strength and weakness that make each of them unique.13

The National Research Council publication Knowing What Students Know14 also makes a plea for increasing the precision of assessment items. That report, also based on recent brain research, and its companion, How People Learn,15 explain that experts display a unique organization of knowledge, and little bits of knowledge are less significant than how an individual organizes information. Assessments that reveal the organization of knowledge within an academic discipline, such as computer-based concept mapping,16 hold great promise for all students.

Universal design approaches to creating online assessments offer opportunities to both refine items and create entirely new kinds of items. Additionally, precise definitions of the constructs being assessed are most likely if a state has well-designed standards that are also outcomes of the universal design process.17 Is the goal to write an essay? To demonstrate knowledge of a paragraph’s structure? To synthesize and summarize information? When the goal is defined, it is easier to ask if it can be demonstrated with unconventional assessment responses, such as a videotaped performance or a conversational description of the correct answer. A universal design initiative in Ohio’s Lorain County, for example, uses skits for formative assessments (i.e., assessments meant to help guide instruction) of students with disabilities.18

mentioned earlier (recognition, strategic, and affective), Dolan and Hall of CAST advocate universal design for large-scale assessment with a system of
- multiple means of recognition
- multiple means of expression
- multiple means of engagement28

The same principles support their work on differentiated instruction through the National Center on Accessing the General Curriculum.29

Guidance also comes from Disabilities Rights Advocates (DRA), which authored Thirteen Core Principles to Ensure Fair Treatment of All Students, Including Those with Learning Disabilities, with Regard to High Stakes Assessments.30 Endorsed by the International Dyslexia Association, National Center for Learning Disabilities, and Learning Disabilities Association of America, these 13 principles focus on policy issues of accommodations, alternate assessments, the IEP, Section 504 of the IDEA, and the appeal process. Additional policy guidance can be found in DRA’s report Do No Harm—High Stakes Testing and Students With Learning Disabilities.31

The National Center for Educational Outcomes (NCEO) has addressed the universal design of assessment in its research.32 The authors of NCEO reports have identified the following as elements of universally designed assessments:
- appropriate for inclusive assessment population
- precisely defined constructs
- accessible, non-biased items
- amenable to accommodations
- simple, clear and intuitive instructions and procedures
- maximum readability and comprehensibility
- maximum legibility33

NCEO has also identified a process for developing universally designed online assessment.34

Step 1. Assemble a group of experts to guide the transformation.

Step 2. Decide how each accommodation will be incorporated into the computer-based test.

Step 3. Consider each accommodation or assessment feature in light of the constructs being tested.

Step 4. Consider the feasibility of incorporating the accommodation into the computer-based test.

Step 5. Consider training implications for staff and students.

**Guidance on Web site (online) accessibility.** Various guidelines for ensuring that Web documents are accessible to individuals with physical disabilities are available.

- The National Center for Accessible Media Web site is ncam.wgbh.org/salt.
- Section 508 of the Rehabilitation Act requires that electronic and information technology used by federal agencies be accessible to people with disabilities, and accessibility standards for 508 compliance are at www.access-board.gov/sec508/guide/1194.22.htm.
- The W3C, an international, nonprofit organization that develops standards and guidelines for authoring documents on the World Wide Web presents its Web Content Accessibility Guidelines (WCAG) at www.w3.org/TR/WCAG10/.
- A number of tools, including the free webxact service (webxact.watchfire.com), check products against these guidelines and standards.
The Vision

Ideally, a universally designed online assessment would allow students to accurately demonstrate their knowledge and skills and educators to receive quick and useful data about their performance. Through instruction and classroom work, students with disabilities will have developed metacognitive skills and discovered tools that best help them understand, remember, and respond. These same tools and strategies would be available to them during assessments. With confidence in the construction of assessment items, the educators could rely on accommodated assessment scores to guide instructional decisions.

One word of caution: An assessment offered exclusively online cannot be considered to be universally designed. Some students have disabilities that prevent them from responding to a computer. A flexible, universally designed assessment environment would therefore offer whichever human, paper, or other interaction works best for each individual.

Recommendations

Based on the discussion in this policy brief, the following recommendations are offered:

• Use the universal design philosophy when first beginning to create an online assessment.
• Include expertise in special education, assistive technologies, and universal design in addition to expertise in assessment, online technology, and subject matter on the team creating the assessment.
• Precisely define the cognitive construct being assessed with each assessment item.
• Consider the three neural networks (recognition, strategic, and affective) when defining the construct being measured.
• Facilitate instructional use of, and increase teacher knowledge of, all assistive technologies that can be used on the assessment.
• Create an environment that encourages individual adjustments in the presentation of and interaction with digital materials.
• Define policy that supports the flexible use of accommodations on an assessment.
• Support research on how accommodations influence the validity and reliability of assessments.
• Encourage technical-standard-making bodies such as IMS (IMS once stood for Instructional Metadata Specifications, but it has since outgrown its acronym) to focus on assessment needs for students with disabilities.

The Kentucky Online Assessment

Four years ago, Kentucky developed an online assessment specifically for students with disabilities. It provides eligible students with disabilities or with limited English proficiency a way to take all portions of the Kentucky Core Content Test online.

The assessment takes advantage of text-to-speech technology: Students (with earphones) can select any text in the assessment and have it read aloud to them. Numbers have not been large enough for reliable data until the 2004-2005 assessment, and the Kentucky Department of Education is awaiting analyses of that information. Meanwhile, anecdotal comments point to success. Representative comments from students include the following:

• “I like being on the computer and not having someone read to me like a kid.” (10th grader)
• “It was easy, and I liked it because I could do it by myself.” (4th grader)
• “It was better than listening to an adult read, and I could go at my own pace.” (7th grader)

The Kentucky Department of Education established widespread instructional use of the text-to-speech technology before using it on assessments. The only students who can use the read-aloud on an assessment are those with IEPs that identify the necessity of the tool. However, any student can use it for schoolwork. That availability has helped many students who process auditory information more easily than they process visual text.
Endnotes


2. Ibid.


7. State policies can be viewed at http://education.umn.edu/nceo/statepolicies.html (Minneapolis, MN: Regents of the University of Minnesota, 2005).

8. Ibid. Examples of accommodations from state assessment policies can be viewed at http://www.ldonline.org/id_indepth/special_education/peer_accommodations.html.


17. See Rose and Meyer, Teaching Every Student in the Digital Age.

18. For more information on the universal design initiative in Ohio’s Lorain County, see http://www.cnet.org/udlnoohio.htm.

19. See Bransford, Brown, and Cocking, How People Learn. This publication encourages specific teaching of metacognitive skills.


23. Ibid.

24. For more information on the National Instructional Materials Accessibility Standard, see http://nimas.cast.org/.

25. For more information on the IMS Question and Text Interoperability specification, visit the IMS Global Learning Consortium Web site at http://www.imsglobal.org/question/.


27. Ibid.


31. Ibid.


