

From: David Gordon [dgordon@cast.org]
Sent: Friday, January 14, 2011 1:08 PM
To: RTTA-RFI
Subject: Assessment RFI Response

Response to the US Department of Education's
Assessment Technology Standards Request for Information (RFI)

January 14, 2011

Submitted By: CAST (Center for Applied Special Technology)
40 Harvard Mills Square, Suite 3, Wakefield, MA 01880; www.cast.org

Contact: David Gordon, Director of Communications,

dgordon@cast.org; tel. 781-245-2212

CAST is pleased to have this opportunity to respond to the Assessment Technology Standards Request for Information. In this memo, we address two questions: Sec. 3.2.1 and Sec. 3.2.28. In addressing the latter, we aim to clarify the meaning of Universal Design for Learning (UDL), which speaks to much more than the very narrow question of accessibility standards for technology products but rather is integral to the whole question of making more effective technology-based assessments that support better teaching and learning.

In the 1990s, CAST introduced the term “universal design for learning,” defined its primary principles, and has since developed, with support from the US Department of Education, comprehensive UDL Guidelines to support its implementation. For more than a decade, CAST also has worked on applying UDL to assessment technology in partnership with leading research organizations, corporations, and state consortia. Thus CAST is especially well positioned to comment on both of these points.

In Section 3.2.1, regarding the current landscape, the following question is asked: “What are the dominant or significant assessment technology standards and platforms (including technologies and approaches for assessment management, delivery, reporting, or other assessment interoperability capabilities)?”

The technical standard which is most promising for supporting diverse learners and providing a full range of accessibility features is the XML-based Accessible, Portable Item Profile (APIP) standard recently

integrated with the Question and Test Item Specification (QTI) and “Access for All” IMS Global Learning Consortium standards.

We are hopeful that APIP will integrate well with Common Cartridge, Learning Tools Interoperability (LTI) and Learning Information Services (LIS). Such integration will eventually provide support for both formative and summative assessments designed to improve teaching and learning. Also significant is the likelihood that matching supports and accommodations provided in instruction with those offered during assessment will improve learning outcomes and the accurate measurement of learning needs and gainsⁱⁱ.

Another promising aspect of the APIP standard is that it will provide metadata to support matching students needs with assessment features without violation of the assessment construct. This will make possible the development of learning and assessment approaches that provide highlights, simplified language, thought questions, graphic organizers, cue items, distraction controls and more to support individual differences within both learning and assessment environments.

In Sec. 3.2.28, the question is raised: “How are issues related to Universal Design for Learning (UDL) relevant to standards for accessible use?”

First, we begin by noting that although UDL is often cited in relation to issues of accessibility for students with disabilities, the principles themselves address the whole population of learners. UDL addresses not only physical, sensory and language barriers (i.e., decoding of text, second-language skills, etc.), but also considers other potential barriers, especially those that are cognitive, executive, and affective. This is important when considering assessment standards because accountability for student outcomes requires one goal above all: the accurate measurement of what students know and can do. Yet research shows that many assessments—in particular many large-scale standardized tests—do not produce a valid and reliable measurement of what significant minorities of students actually know, especially students with disabilities, English language learners, or those from varied cultural backgrounds.

Federal statute now defines UDL as “a scientifically valid framework for guiding educational practice that — (A) provides flexibility in the ways information is presented, in the ways students respond or demonstrate knowledge and skills, and in the ways students are engaged; and (B) reduces barriers in instruction, provides appropriate accommodations, supports, and challenges, and maintains high achievement expectations for all students, including students with disabilities and students who are limited English proficient” (Higher Education Opportunity Act of 2008; *emphasis ours*).

Congress added this definition of UDL to distinguish the concept from universal design (UD), which typically addresses physical and sensory barriers to access but not the other kinds of barriers identified by

recent research in the learning sciences, including neuroscience.

In the same way that many assessments impose physical and sensory demands—demands that are irrelevant to the knowledge and skills actually being measured—the learning sciences reveal that assessments also impose cognitive, executive, and affective demands that are irrelevant. If all students were equal, the “irrelevant” demands of the item would have little importance (e.g., if every student has 20/20 vision, then the visual demands are of little significance because they are the same for everyone). But all students are demonstrably not equal. They differ as much in their underlying cognitive, executive, and affective abilities as in their physical abilities. As a result, fixed or “standardized” items pose very different demands for different students—demands that are easy for some, impediments or barriers for othersⁱⁱⁱ.

What UDL means for large-scale assessments

Applying UDL principles, which take into account the full range of individual differences relevant to learning, affords more accurate measurement by providing options in the assessment instruments—options that reduce “undesirable” difficulties or irrelevant barriers that actually interfere with accurate measurement. This is especially true when UDL principles are applied to the design of assessments delivered with digital technology, which have the potential to provide the kinds of options in the design of items to ameliorate sources of construct irrelevant variance.

To understand the need for options in assessment that support what Congress calls “flexibility in the way information is presented,” consider the following example. English language learners typically start school with far less vocabulary knowledge than those whose first language is English, and this gap may persist throughout the school years. For assessments where particular vocabulary knowledge is not relevant to what is being measured (the causes of the Civil War, for example), providing vocabulary support via a glossary, thesaurus, images, and so forth could enable a more accurate measurement of what the student knows about the relevant content. Absent these options, vocabulary knowledge—which is not being measured—would actually skew the results because the assessment instrument itself relied on a single method of presenting information (the particular vocabulary). Even a student who understood perfectly the causes of the Civil War could be “mismeasured” by not knowing that “garrison” meant the same as “fort.” Offering alternative representations of “garrison” would make that less likely—and strengthen our ability to measure the relevant knowledge.

Options and flexibility are also essential to ensure that the mode of expression or action in the assessment instrument does not introduce unintended, construct irrelevant, barriers. For example, students with certain learning disabilities might require support or flexibility in preparing and organizing a response or

remembering details. Examples of options include providing organizational aids and checklists, or prompts for monitoring time and progress. Without such options, the assessment will likely be inaccurate for such students.

In terms of engagement, the fact that no two individuals approach assessment with the same expectation and motivation raises a concern about how accurate “standardized” assessments are. While it is possible to standardize external conditions, it is not possible to standardize their effects. Some students, for example, are highly engaged by spontaneity and novelty, but others are disengaged or even frightened by those aspects in the environment. As a result, every assessment is to some extent measuring the student’s individual reactions to the conditions of the assessment method, which in turn affects motivation (positively or negatively). Virtually every measurement instrument is inevitably measuring, and therefore confounded by, variations in individual engagement. With the right flexibility and support, such variations can be addressed without risking the validity of the assessment.

Preserving the integrity and validity of any assessment instrument is essential if that instrument is to be useful as a gauge of systemic performance. Only by applying UDL principles (which, by definition, include UD concerns as well as additional barriers that threaten the assessment construct) in a principled way can we ensure that large-scale assessments validly and accurately measure all students’ progress in ways are not unduly influenced by factors that should be irrelevant, such as disability or language barriers.

As one leading expert on educational assessment, Robert J. Mislevy, has stated: “If UDL is applied in a principled manner, it will actually increase construct validity for a larger population of students.”^[iii] (For a full description of the UDL principles, guidelines, and checkpoints, as well as the research basis for each of them, see the UDL Guidelines at the National Center on Universal Design for Learning: www.udlcenter.org/aboutudl/udlguidelines.)

What UDL means for formative assessments

We agree with the [National Education Technology Plan \(2010\)](#) when it states: “When combined with learning systems, technology-based assessments can be used formatively to diagnose and modify the conditions of learning and instructional practices while at the same time determining what students have learned for grading and accountability purposes. Both uses are important, but the former can improve student learning in the moment ...” (p. vii)^[iv]

In addition to improving summative assessments, today’s new technologies make it possible to apply UDL principles to formative assessments that can help teachers detect and correct deficiencies in the curriculum *before* students fail. In fact, UDL asserts that learning, instruction, and assessment are most

effective in environments that are flexible enough to accommodate individuals according to their particular strengths and needs, which may be physical, intellectual, and/or motivational (Rose & Meyer, 2002).

The advantages of computer-based formative assessments for teaching and learning are demonstrated by one OSEP-funded research study in which we applied UDL principles and curriculum-based measurement, a form of formative assessment, in a technology-based learning environment. Both teachers and students responded favorably to the accessibility, immediacy, and computer-based delivery. Teachers reported that having progress monitoring measures that are easily scheduled and administered using the computer software, and include timesaving automatic/semi-automatic scoring, made it possible to have specific information about *each* student enabling them to consider adjustments in instruction. One teacher noted that in some cases the read aloud information revealed that her students were much poorer readers than she realized and these measures allowed her to understand their reading challenges. When we questioned students, they noted how much they liked seeing *specific* and *immediate* results and they enjoyed challenging themselves to do better on the next probe. Because most students are familiar with on-demand results when they use computers, they expected and received immediate results from the maze selection measures, for example, and liked seeing scores and comparing results over time.

Ultimately, CAST envisions a day when performance-based accountability will be measured by assessments that take place much closer to the instructional episode so that they can be used to improve academic performance *before* some learners fail as well as provide fair and appropriate accountability data.

Public policy (i.e., the Race to the Top competition) has already recognized the need for “hybrid” large-scale assessments that support a culture of continuous improvement at the school-building level by providing useable data for instruction while also supporting accountability. Already, progress monitoring—the scientifically-based practice that is used to assess students' academic performance and evaluate the effectiveness of instruction—can be implemented with individual students or an entire class. Progress monitoring can be used to shape instruction to help move students toward meeting state standards while also providing data that can be used to determine adequate yearly progress. [\[v\]](#)

Embedding continuous assessment in instructional materials and methods themselves through the kind of technology-rich, UDL-based curriculum recommended by the National Educational Technology Plan would make it possible to assess not only students and their teachers but the curriculum itself. This would allow the collection of voluminous and timely data on the effectiveness of every element in the curriculum: what works, what doesn't work, and what works for whom. The result: comprehensive accountability systems and instructional reforms that could support robust learning opportunities for all.

-
-
- [i] Luke, S.D., & Schwartz, A. (2007). Assessment and accommodations. *Evidence for Education II*(1): 4. Retrieved Jan. 13, 2011 from http://www.nichcy.org/Research/EvidenceForEducation/Documents/NICHCY_EE_Accommodations.pdf
- [ii] Rose, D.H. & Vue, G. (2010). 2020's Learning Landscape: A Retrospective on Dyslexia. *International Dyslexia Association, Perspectives on Language and Literacy*, 36(1), 33-37.
- [iii] Gordon, D.T., Gravel, J.W., & Schifter, L.A. Perspectives on UDL and assessment: An interview with Robert J. Mislevy, in D.T. Gordon, J.W. Gravel, & L.A. Schifter (Eds). *A policy reader in Universal Design for Learning* (pp. 209-218). Cambridge, MA: Harvard Education Press.
- [iv] US Department of Education, Office of Educational Technology (2010). Transforming American education: Learning powered by technology. Draft National Educational Technology Plan 2010. Washington, DC: Author.
- [v] Fuchs, L.S., & Fuchs, D. (2007). Determining adequate yearly progress from kindergarten through grade 6 with curriculum-based measurement. LDOnline. Retrieved Jan. 11, 2011 from <http://www.ldonline.org/article/14601>