

Perspectives on Large-Scale Assessment, Universal Design, and Universal Design for Learning

July 7, 2010

Introduction

In recent decades, the aspiration that all children can and should learn to high achievement standards has won support in federal legislation (NCLB/ESEA, 2002; IDEA, 1997, 2004) and civil rights statutes (ADA and Section 504). As a result, all states are required to establish large-scale assessment programs to test the progress of all students toward meeting the standards. These assessments can have significant implications for states, districts, and schools, as well as for individual teachers and students, as public policies are shaped by their results. Even local economies and property values can be affected by test scores. Because of these far-reaching effects, accountability for student outcomes requires one goal above all: the accurate measurement of what students know and can do.

The challenge facing policymakers then is to endorse, fund, and recognize assessment regimes that accomplish this goal. Yet many questions have been raised about the current state of large-scale assessments for accountability purposes—in particular whether these standardized tests fail to 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. Without accurate measurement, accountability systems are not only ineffective, they are unethical.

In recent years, CAST has worked with a number of organizations and agencies across the United States to explore ways of applying universal design (UD) and universal design for learning (UDL) principles to the development of large-scale assessments. In this statement, CAST aims to clarify the implications of UD and UDL for assessments that are used for accountability purposes.

Definition of Terms

Universal design, a concept pioneered by architect and disability-rights advocate Ron Mace in the 1990s, is defined by federal statute 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 interoperable with assistive technologies” (Assistive Technology Act of 1998). In 1997, the Center for Universal Design at North Carolina State University developed a set of seven general principles to guide such designs: 1) equitable use; 2) flexibility in use; 3) simple and intuitive use; 4) perceptible information; 5) tolerance for error; 6) low physical effort; and 7) size and space for approach and use.¹

Around this same time, researchers at CAST and colleagues in the field looked for ways to apply UD principles to education. Recognizing that principles used to design things (such as buildings and products) are not necessarily sufficient for the design of human interactions (such as teaching and learning), CAST and colleagues instead looked to new discoveries in science and technology for insights into how to create more flexible, supportive, and inclusive curriculum. Drawing on this research, CAST articulated a new set of principles based not in architecture and product design but in the learning sciences, including cognitive neuroscience—the principles of universal design for learning (UDL), a term first used by CAST as early as 1995.

UDL requires that we not only make information accessible—which, after all, is just one component of learning—but that we also support learning activities, expression, and engagement. To do this, CAST identifies three main principles of UDL:

1. *Multiple means of representation* to give learners various ways of acquiring information and knowledge,
2. *Multiple means of action and expression* to provide learners alternatives for demonstrating what they know, and
3. *Multiple means of engagement* to tap into learners' interests, challenge them appropriately, and motivate them to learn² (Rose & Meyer, 2002).

UDL was defined by federal statute in 2008 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).

Applying Universal Design to Assessment

For more than a decade, the National Center on Educational Outcomes at the University of Minnesota has led efforts to apply the seven UD principles to large-scale assessment. These assessments are designed and developed from the outset to minimize the effects of disability, race, culture, gender or English language ability on testing while still providing valid inferences about performance for all students who participate in the assessment.

Rather than retrofitting assessment instruments that are inaccessible to students with disabilities or other learning challenges, UD-based large-scale assessments aim to reduce the need for accommodations and alternative assessments by eliminating access barriers when the tests are made. In some cases, computer-based testing has proven to be a promising aid to the universal design and delivery of assessments³.

The primary goal of UD in large-scale assessment is to ensure accurate measurement of academic performance for all participating students (Johnstone, 2003; Thompson & Thurlow, 2002). In a climate where educational policy has placed increasing emphasis on the use of large-scale assessment to assess school performance for accountability purposes, the application of UD principles to assessment has been an important advance, providing students with greater opportunities to demonstrate their proficiency.

UDL and Large-Scale Assessment (LSA)

CAST and its partners have also worked to apply UDL principles to large-scale assessments. What does UDL add that is not already addressed by universal design practices? Because UDL is based in the modern learning sciences rather than architectural practices, the UDL Guidelines draw attention to a broader set of barriers that are relevant to accurate measurement, specifically the measurement of learning. Where UD focuses

primarily on physical, sensory and some language barriers (i.e., decoding of text, second-language skills, etc.), UDL broadens the scope considerably to include other potential barriers, especially those that are cognitive, executive, and affective.

In the same way that almost all items impose physical and sensory demands—demands that are irrelevant to the knowledge and skills actually being measured—modern learning sciences reveal that almost all items also impose cognitive, executive, and affective demands that are also irrelevant. If all students were equal, then 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 others.

Applying UDL principles, which take into account the full range of individual differences relevant to learning, allows more accurate measurement by providing options in the assessment instruments—options that reduce “undesirable” difficulties or irrelevant barriers that actually interfere with accurate measurement. Take the example of English language learners, who typically start school with far less vocabulary knowledge than those whose first language is English. This knowledge 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 traditional large-scale assessment instrument is essential if that instrument is to be useful as a gauge of systemic performance. Indeed, the No Child Left Behind Act of 2002 mandates:

“§200.2 State responsibilities for assessments—

The assessment system required under this section must meet the following requirements:

... Be designed to be valid and accessible for use by the widest possible range of students, including students with disabilities and students with limited English proficiency.“ [emphasis ours]

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.”⁴

(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.)

The Future of Assessment for Accountability

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)⁵

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.⁶

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.

¹ Center for Universal Design (1997). *The principles of universal design, version 2.0*. Raleigh, NC: North Carolina State University.

² Rose, D.H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. Alexandria, VA: ASCD.

³ Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). [*Universal design applied to large scale assessments*](#) (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

⁴ 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.

⁵ 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.

⁶ Fuchs, L.S., & Fuchs, D. (2007). Determining adequate yearly progress from kindergarten through grade 6 with curriculum-based measurement. LDOnline. Retrieved June 10, 2010 from <http://www.ldonline.org/article/14601>