

Rethinking Basic Assumptions of Test Development: Assessment Frameworks for Inclusive Accountability Tests

Background

For the *No Child Left Behind Act of 2001* (NCLB), states are required to assess the achievement of all students on state-defined grade level content standards. They are to use the results to hold schools, and the state, accountable for all students achieving state-defined proficiency on grade-level content by the year 2014. Schools and the state are to show “adequate yearly progress” (AYP) toward reaching that goal in the meantime. Results from each year’s assessment of student achievement can be used to redesign instructional programs so that more students are successful each year.

If educators were asked today whether all students are achieving at the proficiency level expected by 2014, virtually no educator in the U.S. could say yes. Some would say that the goal is unrealistic; others would say that fundamental changes in instructional approaches, structures, and

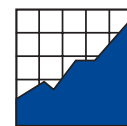
budgets have to occur to make the goal achievable.

Some educators see a need to improve assessments so that they inform instruction on grade level content. These educators are calling for assessments based on a limited number of clearly defined and important constructs that reflect what all students should know and be able to do at grade level, and that are constructed to yield unambiguous evidence of whether groups or classrooms of students have mastered those constructs at grade level. They are rethinking common assumptions about the entire test development process—from the definition of the constructs to be tested, through development, to implementation.

This requires a transparent discussion on how to put into practice the new assumptions about teaching and learning that have come out of standards-based reform so that assessments can be designed to meet the

accountability requirements of NCLB, yet give better information about the performance of students within the prescribed content. Compatible classroom assessments can provide additional diagnostic information for all students. Clarification and articulation of content standards—that is, precisely defined constructs—are the foundation of good assessments in order to maximize the measurement of achievement rather than the effects of disability, cultural background, or socioeconomic status.

As an example, the Commission on Instructionally Supportive Assessment (2002) identified nine requirements for assessment



NATIONAL CENTER ON EDUCATIONAL OUTCOMES



The College of Education & Human Development

UNIVERSITY OF MINNESOTA

Table 1. Instructionally Supportive Assessments

Purpose, Procedure, and Challenges
<p>Purpose: To rethink test development procedures, starting from the alignment of content standards, instruction, constructs being assessed, and assessment. Instructionally supportive assessments are based on the belief that required accountability tests must be useful to educators concerned about improving the instruction of children.</p> <p>Procedure: Starting from a very comprehensive approach, the Commission on Instructionally Supportive Assessment (2002) identified nine requirements for assessment development:</p> <ol style="list-style-type: none">1. Prioritized content standards.2. High-priority content standards described.3. Standard-by-standard reporting for high-priority standards.4. Optional classroom assessment procedures for other standards.5. Breadth of curriculum monitored by the state.6. Well designed assessments with accommodations and alternate methods of assessment.7. Minimum of 3 years to produce assessments.8. Professional development on how to optimize children’s learning.9. Evidence of appropriateness for accountability, assessment of standards, and enhancing instruction, and lack of negative consequences. <p>Specific reasons to support each of these requirements are provided. The requirements can be met either by drawing on the capabilities of state agencies or by issuing competitive requests for proposals to firms or individuals capable of carrying out one or more of the required activities.</p> <p>Challenges: Instructionally supportive assessment involves rethinking the entire system, including the content standards, aligned curriculum, and rich and varied instruction, to ensure that educators know what to teach and that children are taught. It is easier to focus on narrow aspects of assessments.</p>

development (see Table 1). These requirements were ones the Commission believed that, if implemented, would create responsible assessments for the improvement of students’ learning.

Opportunities for Students with Disabilities

Dr. Jim Popham, chair of the Commission on Instructionally Supportive Assessment, summarized what he considers to be the implications of the assessment requirements of the *No Child Left Behind Act* for children with disabilities. This summary is presented in Table 2.

There are opportunities created when “extremely significant” content standards are identified, as explained in point 6 in Table 2. For example, universally designed assessments, which are assessments designed from the beginning to be accessible and valid for the widest range of students, are typically portrayed as one aspect of assessments that are designed to support instruction and accountability for all students. The identification of a small number of “extremely significant” content standards to be the focus of large-scale assessments in itself can be viewed as an aspect of a universally designed assessment.

Benefits of merging these ideas for students with disabilities are shown in Table 3.

Opportunities and Implications for State Policymakers

As state policymakers grapple with the challenge of truly “rethinking” entrenched approaches to content standards and assessment, there are practical and political issues to consider. These include the “yeah, but...” scenarios heard by assessment directors in discussions among policymakers and practitioners across the country. Here are a few

Table 2. Inclusive Assessment Requirements and Instructionally Supportive Assessments

Implications Identified by Jim Popham

1. Federal law now requires that essentially all American children, including those with disabilities, must complete state assessments (sometimes with disability-specific accommodations and alternate assessments) in seven grade levels.
2. Irrespective of whether one regards such assessments as educationally desirable, it is now a federal requirement.
3. As is often the case with federal or state education-related laws, such statutes should be implemented in a manner that is most educationally beneficial to the children affected by those laws.
4. In order for state assessments to be educationally beneficial, they must possess two assessment attributes, namely, they must (a) describe with clarity the skills and/or knowledge they assesses and (b) provide results in a form so that teachers can identify which parts of their test-related instruction were effective or ineffective.
5. These two assessment attributes *cannot* be satisfied by standardized tests that are (a) traditionally constructed to permit norm-referenced interpretations, or (b) customized to provide standards-based interpretations about students' mastery of enormous numbers of curricular outcomes.
6. Thus, to optimize the educational benefits for all children who must complete the new federally required achievement tests, including those children with disabilities, statewide tests intended to satisfy the new federal requirements must attempt to assess only a small number of *extremely significant content standards*.

(See NCEO teleconference materials at <http://education.umn.edu/nceo/Presentations/tele5.htm>)

of the scenarios and responses related to the implications for students with disabilities of the NCLB assessment requirements.

► **Yeah, but how can we expect essentially all students, who may be instructed at very different levels, to participate in the same tests?**

Response: Federal laws require not only assessments, but also require identical content expectations for all students (other than the small exception proposed for students with significant cognitive disabilities). NCLB requires specifically that all children be assessed on grade level content and achievement standards. There are multiple solutions to this “yeah but...” and they include revisiting

how we design instruction and curriculum to ensure all students learn the challenging content from the earliest grade levels.

Clarification and articulation of extremely significant content standards is essential to ensuring a coherent, aligned, and focused teaching and learning system that is accessible to all learners. Even with shifts in how we ensure all children learn, we still have to address the adequacy of current assessment approaches. With the opportunities of instructionally supportive and universally designed assessments aligned to extremely significant constructs, many more students can meet the assessment targets.

► **Yeah, but if we hunker down, this too shall pass.**

Response: The basic assumptions of standards-based reform that underpin the Federal laws – all students, meeting high standards, and accountability on the part of schools and states – have been in place for over a decade. These assumptions have broad bipartisan support, and have shown remarkable political and popular durability. Also, these assumptions are consistent with the predominant agenda of public educators at local and state levels, that is, of meaningful education with high expectations for all students. Looking for ways that we can use the law to further our agenda of improving results for all students, and improving

Table 3. Benefits of Instructionally Supportive Assessments for Students with Disabilities

Benefits
<ol style="list-style-type: none"> 1. By identifying and clearly defining a reasonable number of content standards, or clusters of content, that are of high importance and challenging, it becomes easier for educators and IEP teams to understand what they need to hold inviolate as they consider and prioritize the focus of a student's learning time. 2. Clear definitions of high importance content standards will also make it easier to see the link from grade to grade, so that students will not miss content needed for higher grades simply because educators and IEP teams did not know of its importance when the student was in an earlier grade. 3. When we have clear descriptions of the high importance content standards, it will be easier to describe what accommodations are appropriate in instruction and in assessment. 4. When the high importance content is identified, it should become clearer how to define and develop the needed research base for accommodations. That is because assessments of higher-level complex learning are less likely to focus on numerous little skills measured by items that may test the very skills that a disability blocks (e.g., decoding, even at upper grade levels). We may find that students with disabilities perform at higher levels when the assessments truly focus on higher-level knowledge and skills.

the quality of our educational assessments to measure the results of instruction more accurately, is consistent with that agenda.

► **Yeah, but how can we change our content standards? Our content experts worry that “if it is not tested, it will not get taught.” And we don’t want to inadvertently narrow or “dumb down” the curriculum.**

Response: This is perhaps the biggest challenge. In many states, content experts and teachers worked over a period of years to negotiate the academic content standards. In that process, they tended to include more content, in part to avoid conflict and to satisfy multiple constituencies in an inherently politicized system. In many states, this may have resulted in a vast number of content standards or benchmarks, too many to assess in the time available for state-level testing

and, perhaps, too many to teach and learn in the time available for instruction.

The pressures created by NCLB provide an opportunity to revisit this approach. Content experts and teachers will agree that schools and teachers should be held accountable for reasonable learning targets, and assessments designed to measure the targets should be reasonable and fair as well. As long as the state-defined content cannot be assessed in total on an assessment, and states have not defined and communicated the content to be assessed, assessment becomes a punishing variation of Russian roulette, with teachers and schools guessing at what will be assessed. No content experts or teachers support that approach, which may be the current reality.

One of the foundations for instructionally supportive accountability tests is that they focus on a modest number

of assessment targets. If the accountability purpose of the test is to encourage teaching of important content, then those targets must be truly significant and communicated clearly to students, teachers, and the general citizenry.

The entire system of state content standards does not have to be redesigned. Instead, states can work to define an *assessment framework* that functions as a companion document to already approved content standards. This requires alignment justifications demonstrating how assessed skills and bodies of knowledge are derived from the state's existing content standards. The identified content clusters and important constructs that are assessed will incorporate many of the smaller grain size content standards or benchmarks.

There is no magic way to do this, but there are some general

principles that should be followed, based on an example used in one state.

The steps that state took can be summarized more generically as:

- Raise awareness and develop consensus among various groups of the need to prioritize or rework content standards.
- For each content area and grade, form groups to identify:
 - o non-essential knowledge and skills;
 - o unnecessary duplications of content across or within grade levels;
 - o content that cannot be assessed through large-scale methods; and
 - o content that can be logically clustered into larger constructs incorporating small grain content.

Share the resulting revisions across content areas and grades as a check on progression and importance.

- Have people who are very familiar with the content area and people less familiar review the resulting content and progression to ensure that the resulting descriptions are considered important by people other than the writing team, and by those who are not content experts.
- Develop an assessment framework based on the resulting content descriptions.

This process does not require new content standards, and specifically avoids identifying “easier” content standards. It instead requires identifying high import challenging content clusters that incorporate essential bits of content and knowledge across the

discipline and across grades. It should result in a limited number of challenging and important constructs that will serve as the framework for the accountability test.

The process could also be modified by using a version of Norman Webb’s *Balance of Representation* protocol. This is another tool that would push people to prioritize the required knowledge and skills for their content areas (see N.L Webb’s *Alignment of Science and Mathematics Standards...* and *Criteria for Alignment of Expectations...* in the Resources).

► **Yeah, but the NCLB accountability requirements are so challenging that many of our schools will fail to meet AYP requirements within the first few years. This seems like an undoable challenge!**

Response: We need to use the best research in teaching, learning, and assessment to help states design assessment systems that can promote student learning. As we organize fewer, high import, challenging content clusters to assess, we also provide a clearer “path” to the essential skills and knowledge. This will allow all teachers and all IEP teams, for the first time in many states, to understand how all the hundreds of “tiny grain size” standards work together to result in durable and important skills and concepts along the grade levels. Although NCLB *accountability* provisions may not allow for a great deal of innovation and tailoring, working to make the *assessment* instructionally sensitive and

universally designed will make a big difference ensuring that all schools, districts, and states are successful. Most important of all, it will make a difference in ensuring that all students are successful.

► Resources

Alignment of Science and Mathematics Standards and Assessments in Four States (NISE Monograph No. 18). Webb, N.L. (1999). Published jointly by the National Institute for Science Education and the Council of Chief State School Officers.

Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education (Monograph No. 6). Webb, N.L. Madison: Wisconsin Center for Education Research, Council of Chief State School Officers and National Institute for Science Education.

Building Tests to Support Instruction and Accountability: A Guide for Policymakers. Commission on Instructionally Supportive Assessment. (2001). AASA, NAESP, NASSP, NEA, NMSA. Available from the American Association of School Administrators Web site at http://aasa.org/issues_and_insights/assessment/Building_Tests.pdf. Also see companion piece on *Illustrative Language for an RFP to Build Tests to Support Instruction and Accountability*, at http://aasa.org/issues_and_insights/assessment/Illustrative_Language_for_an_RFP.pdf.

Crafting Curricular Aims for Instructionally Supportive Assessment, with Examples of Appropriate Curricular Aims.

Popham, W. J., Farr, R., & Lindquist, M. (2003). Available from the National Center for Educational Outcomes at <http://education.umn.edu/NCEO/Presentations/CraftingCurricula.pdf>.

Materials for the January 27, 2003 NCEO teleconference: Part Two of "Building Tests to Support Instruction and Accountability for All Students."

Popham, W. J., Thurlow, M.L., & Marion, S., Presenters. (2003). Available from the National Center on Educational Outcomes at <http://education.umn.edu/nceo/Presentations/tele5.htm>.

Universal Design Applied to Large-Scale Assessments

(Synthesis Report 44). Thompson, S.J., Johnstone, C. J., & Thurlow, M. L. (2002). Minneapolis, MN: University of Minnesota, available from the National Center on Educational Outcomes at <http://education.umn.edu/nceo/OnlinePubs/Synthesis44.html>.

Universally Designed Assessments: Better Tests for Everyone! (Policy Directions 14). Thompson, S., & Thurlow, M. (2002). Minneapolis, MN: University of Minnesota, available from the National Center on Educational Outcomes at <http://education.umn.edu/nceo/OnlinePubs/Policy14.htm>. ▲

About NCEO

The National Center on Educational Outcomes (NCEO) was established in 1990 to provide national leadership in the identification of outcomes and indicators to monitor educational results for all students, including students with disabilities. NCEO addresses the participation of students with disabilities in national and state assessments, standards-setting efforts, and graduation requirements.

The Center represents a collaborative effort of the University of Minnesota, the Council of Chief State School Officers (CCSSO), and the National Association of State Directors of Special Education (NASDSE).

The Center is supported primarily through a Cooperative Agreement (#H326G000001) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. Additional support for targeted projects, including those on LEP students, is provided by other federal and state agencies. The Center is affiliated with the Institute on Community Integration in the College of Education and Human Development, University of Minnesota. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.

NCEO Staff

Deb A. Albus
Michael Anderson
Ann T. Clapper
Jane L. Krentz
Kristi K. Liu
Jane E. Minnema
Ross Moen
Michael L. Moore
Rachel F. Quenemoen
Dorene L. Scott
Sandra J. Thompson

Martha L. Thurlow, director

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

NCEO *Policy Directions* is a series of reports that address national policy issues related to students with disabilities. This report was prepared by Rachel Quenemoen and Scott Marion. It is available in alternative formats upon request.

Additional copies may be ordered for \$3.50. Please write:

Publications Office
NCEO
350 Elliott Hall
75 East River Road
University of Minnesota
Minneapolis, MN 55455
Phone: 612-624-8561
Fax: 612-624-0879

Visit our Web site at:
<http://education.umn.edu/NCEO>