Technical Considerations for Developing Assessments That Include Special Populations and Are Based on Organized Learning Models

Topic 3 White Paper

P. Almond N. Kingston, H. Michaels, E. Roeber, S. Warren, P. Winter, and C. Mark

Presented at the
Colloquium on Learning Models (Learning Maps/Progressions/etc.), Instruction, and Next Generation
Assessments that Include Special Populations
October 26, 2012
Washington, D.C.

Acknowledgments

Audience and Purpose

Target audiences for this paper are researchers, cognitive scientists, instructional design experts, test developers, and practitioners. Researchers with expertise in the cognitive sciences will be needed to help increase understanding of cognitive pathways. Instructional design experts can help translate what is known about the cognitive pathways into learning progressions. Instructional experts can help design learning environments that maximize the effective implementation of learning progressions. Test developers will need to draw on research findings and the expertise of cognitive scientists, technology and instructional designers, psychometricians, and engineers to develop reliable and valid assessments. Ultimately, practitioners who work with the students will need expertise in diagnosis, intervention, and assessment. Clearly, a teamwork approach is required to realize the potential of the research agendas proposed in the papers to strengthen the validity of next-generation assessments that are appropriately inclusive of special student populations.

SRI/CETE Invitational Symposium

Using Cognitive Learning Models to Inform the Development of Assessments for Students in Special Populations

Symposium Participants

Sponsors   Renée Cameto, SRI International
             Neal Kingston, CETE, University of Kansas

Participants   Topic 1

             Karin Hess: Facilitator, NCIEA
             Sue Bechard: Co-facilitator, IEA
             Rosemary Abell, R.O.S.E. Consulting
             Diane Browder, UNC Charlotte
             Chris Camacho, WestEd
             Leah McConaughey, NY DOE
             Michael Russell, Measured Progress
             Kelli Thomas, University of Kansas

             Topic 2

             Edynn Sato: Facilitator, WestEd
             Renée Cameto: Co-facilitator, SRI International
             Gary Cook, University of Wisconsin
             Danielle Griswold, Achieve
             Nancy Harayama, Boston University
             Donna Lehr, Boston University
             Kate Nagel, SRI International
             Alan Sheinker, CETE, Kansas University

             Topic 3

             Neal Kingston, Facilitator, CETE, University of Kansas
             Patricia Almond: Co-facilitator, CATE University of Oregon
             Carrie Mark, University of Kansas
             Hillary Michaels, WestEd
             Edward Roeber, Michigan State University
             Sandra Warren, ASES SCASS
             Susan Weigert, OSEP, US DOE
             Phoebe Winter, Pacific Metrics

Abstract

A symposium on understanding learning progressions and learning maps to inform the development of assessments for students in special populations was held in July 2011. The symposium posed the following question for work group 3's consideration: "What technical considerations are important for developing learning progressions or learning maps and employing them to develop and implement comprehensive assessment systems that include special populations?" This white paper reports the substance of the work group's deliberations.

For special populations (e.g., students with disabilities, English language learners [ELLs]), principles like universal design, evidence-centered design, and instructional scaffolding now inform standards, assessment, and instruction. In this regard, the symposium addressed organized learning models (OLMs), comprehensive and inclusive assessment systems, and special student populations. Technical considerations regarding OLMs and the assessment systems developed on the basis of those OLMs entail several areas of validity, particularly the validity of the OLMs themselves and the validity of inferences derived from assessments based on the OLMs. A principal expectation in developing, validating, and employing OLMs to design assessment systems is that their implementation will result in improved instruction and learning. Work group deliberations raised a number of questions. Will the expected improvement in instruction and learning ensue from the use of OLMs? What is the evidence for change? How should we interpret that evidence to evaluate the outcomes from using OLMs? What will be the overall effect on students and educational institutions? These questions provide the core focus for this white paper.

Introduction

Focus on universal design principles in standards, instruction, and assessment can result in enhanced accessibility for expanded groups of users (Thompson, Johnstone, Anderson, & Miller, 2005). Likewise, specific focus on diverse groups of students can lead to a system of OLMs (e.g., learning progressions [LPs]—in science, learning trajectories—in mathematics, developmental continuums—in reading, or learning maps [LMs]) that is accessible to the broadest group of learners. Such focus may result in comprehensive assessment and accountability systems—ones that are based on professionally established standards, relevant to the widest range of students, and sensitive to individual needs and student performance throughout the assessment and reporting process.

Successfully employing OLMs in the development and implementation of education systems (instruction and assessment) that include special populations of students is likely to result in education systems that work for all students, including special populations. To this end, the work group 3 discussed Technical Considerations, while our symposium colleagues discussed "Using Learning Progressions and Learning Maps as a Foundation for New Assessments" (Topic 1), and "Alternate and Multiple Pathways for Special Populations" (Topic 2). This "technical considerations" paper builds on Topic 1 and Topic 2 findings and conclusions.

For the varied terms that surfaced on the first day of the symposium—LPs, LMs, learning trajectories, learning progress models—the work group 3 sought a common, encompassing term. For our symposium discussion, we settled on OLM; that is, learning sequences that vary in grain size, complexity, and chain length. Note, however, that consensus about an encompassing term was not achieved across all work groups. In this paper we employ OLM as the term for reporting work group 3's observations and queries.

Four consortia (two General Supervision Enhancement Grants [GSEGs] and two Race to the Top [RttT] programs) are developing next-generation LPs or LMs linked to the Common Core State Standards (CCSS) for an educational system intended to prepare students for college and career readiness. The assessments the four consortia are developing aim to monitor growth and measure achievement in relationship to these LPs/LMs for accountability purposes. Much of the current work addresses typical sequences that students traverse in mastering knowledge in a given content area. When systems include special populations, how existing knowledge, theory, and research concerning how LPs and LMs have been developed and employed needs to be considered in designing and developing assessments. That understanding also needs to be extended to the learning and achievement of special population students as they achieve mastery. Expanded understanding of LPs and LMs needs to address their potential effects on assessment and instruction for special populations. In addition, the development of technically defensible assessments for these populations needs to be considered.

OLMs and Comprehensive Assessment Systems

The first white paper in this series addresses the question: "Why cognitive learning models should be used as the foundation for designing next generation assessment systems?" (Bechard, Hess, Camacho, Russell, & Thomas, 2012). That papers' authors point out that a quality, comprehensive assessment system should strategically balance summative, interim, and formative assessments, with each component designed for a particular purpose and use. They highlight how the use of cognitive learning models as a foundation for next-generation assessment systems may improve on current status and achievement-based accountability measures. They predict that these new comprehensive assessment systems, founded on cognitive learning models, will be: more sensitive to the performance of students at the highest and lowest levels of achievement, describe changes in student knowledge and skills over time, improve the utility of assessment results for teaching and learning, and provide a framework for teachers' tracking of student performance and for supporting their decisions for adapting instruction to their students' individual needs. The first white paper focuses on how teaching and learning could be influenced by assessment systems based on cognitive learning models. Throughout, the authors refer to an assessment system that is instructionally sensitive and provides information that improves teaching and learning.

The first white paper hypothesizes that founding comprehensive assessment systems on cognitive learning models (e.g., learning progressions—in science, learning trajectories—in mathematics, developmental continuums—in reading, or LMs) will improve the utility of assessment results and inform instructional decisions better. The paper presents several cognitive learning models as examples, discusses variations in their levels of detail or granularity, and sets forth an ongoing iterative process involving development based on research, expert reviews, and collection of empirical data to validate the models.

The first white paper notes that comprehensive assessment systems pose a challenge. Results from formative, interim, and summative assessments have not, in the past, been interpretable across different levels of the education system. Moreover, summative assessment data do not easily scale down to the classroom, and formative assessment data do not easily scale up to the district/state level. The authors conclude that, although building next-generation assessment systems based on a cohesive learning model framework is desirable, when the systems are implemented, the interpretations of student progress that teachers use in their classrooms and those that underlie the designs of large-scale summative assessments must cohere.

Systems that support improved teaching and learning serve to expand our understanding of their validity, by emphasizing inferences from assessment results designed to inform (formative) instruction at the individual student and classroom levels. In a preliminary theory of action for Cognitively Based Assessment of, for, and as Learning (CBAL), Bennett (2012) reframes the modern validity argument by combining a theory of action with CBAL's research program. He gives greater prominence to the effects of the assessment system on individuals and institutions than to technical quality. Bennett emphasizes both instrument technical quality and system consequences, as well as the usefulness of results for improving teaching and learning.

His greater prominence on effects of the assessment system seems to fit well with three of Darling-Hammond's (2010) critical characteristics of high-quality comprehensive assessment systems: involving teachers in the development and scoring of assessments, structuring assessments to support improved teaching and learning, and ensuring that assessment systems provide useful information for all stakeholders (see the first white paper).

To ensure that OLMs apply to students in special populations focus on the specific characteristics of those students and on the principles of universal design is needed during the iterative development of OLMs. Inclusion of special populations from the outset will ensure that assessments based on the OLMs apply to all students.

Comprehensive assessments include formative, interim, and summative components, yet much of the interest in OLMs centers on formative assessments. Using OLMs as a basis for developing comprehensive assessment systems thus holds promise for measuring growth and monitoring progress for students in special populations. One challenge for comprehensive systems is that, until now, summative assessments have not provided suitable information to support classroom-based instructional decisions for most students, special populations in particular. At the same time, formative assessment results have not been used to aggregate scores for accountability purposes.

Technical Considerations

Legitimating Learning Models and Validating Assessment Systems

Work group 3 focused on the question, "What technical considerations are important for developing learning models and employing them to develop and implement comprehensive assessment systems that include special populations?"

On one hand, technical considerations in relation to assessments have historically addressed reliability and the validity of inferences based on assessments in the tradition of the *Standards*

for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). The *Standards* outline considerations for evaluating the validity of interpretations on the basis of assessment results. However, in its February 2010 Policy Brief on *Principles for a Comprehensive Assessment System*, the Alliance for Excellent Education (AEE) proposed that, "A new assessment system would redesign the summative tests used for accountability purposes and embed them in a comprehensive and coherent system in which curriculum, instruction, and assessment are intertwined." We may thus need to determine whether our current practices for evaluating and documenting assessment validity should be enhanced to meet the multiple purposes and multiple levels of information these new comprehensive assessment systems require.

On the other hand, technical considerations associated with learning models refer to the process of development and the "iterative" process Bechard et al. (2012) described. Those processes start by selecting a construct, then identifying student understandings and misconceptions of the construct based on a review of research, with grouping of ideas into sets. The resulting hypothesis about this progression is then revised in response to new information. Next, assessment items are developed and administered. Data from the assessment may lead to new ideas or a reorganization of the progression. According to Alonzo and Steedle (2008), however, longitudinal studies are required to fully validate a learning model.

Technical areas discussed below include legitimating OLMs and associated technical issues that relate to making inferences that inform teaching and learning for students in special populations. We also address reliability and validity as they relate to assessment systems for those populations. Future research will need to address these issues.

Technical considerations for OLMs

Developing OLMs and legitimating learning models that include special populations entail several technical considerations.

In describing an iterative process for developing a learning progression in science with associated assessment items, Alonzo and Steedle point out that:

> Learning progressions have been proposed for use in both large-scale and classroom assessments. In both cases, they may provide more detailed information about student thinking than more traditional models of assessment. This detailed information is particularly important in the classroom, where it can be used as the first step in a formative assessment process, to impact instructional decisions and provide feedback to students, ultimately improving student learning (Alonzo & Steedle, 2008, p. 419).

In assessing studies designed to diagnose students' learning progression levels, they note that, although the language used in the items seemed to reflect student thinking, misinterpretation of that language may lead to inaccurate diagnoses for a subset of students. They conclude that identified issues are less problematic for classroom applications than for learning progressions in large-scale testing. For work group 3, this distinction between applying learning progressions to formative assessment versus using them as a foundation for large-scale assessment posed a conundrum. Alonzo & Steedle conclude that:

> As predicted, students [did not] respond consistently to similar problems set in different contexts. Although the language used in OMC items generally seems to reflect student thinking, misinterpretation of the language in items may lead to inaccurate diagnoses for a subset of students. Both issues are less problematic for classroom applications than for use of learning progressions in large-scale testing (Alonzo & Steedle, 2008, p. 389).

Understanding how and why "similar problems set in difference contexts . . . . may lead to inaccurate diagnoses for a subset of students" informed the work group concerns about making valid inferences for students in special populations.

Technical considerations for inclusive assessments founded on OLMs

Although published more than 12 years ago, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) continues to serve as the basis for determining the technical adequacy of assessments. Recent education initiatives resulting from the Elementary and Secondary Education Act (ESEA) and RttT priorities call for a new generation of assessments based on CCSS. These assessment frameworks may include some OLM components.

Although technical considerations regarding assessments development based on OLMs can apply to the complete range of students in prekindergarten through grade 12, the nuances relevant to students who develop knowledge and skills differently from the "typical" student may be overlooked. Those groups of students include students with Individualized Education Programs (IEPs), those with 504 plans, those considered to be persistently low performing, ELLs, and those considered to be talented and/or academically gifted. We address critical issues related to targeted groups of students to whom OLMs as typically presented may not apply.

As the Topic 1 White Paper discusses, using OLMs in developing formative assessments and influencing classroom instruction is relatively straightforward. OLMs can be employed to design and develop the summative assessments ESEA requires for current accountability, particularly if the summative assessments are fixed forms used to assess all students. Assessments whose results measure performance progression from novice to master can indicate how instruction could be changed.

Drawing on its expertise in OLMs, special populations, and assessment, the group looked at the problem from three perspectives: (1) using assessment results to inform instructional decision-making at the student and classroom level; (2) monitoring progress and addressing academic achievement and accountability at the district and state levels; and (3) determining how OLMs relate to special population students' learning and academic development.

Summary

Although, the field of educational measurement has evaluated the validity of inferences based on assessment results for decades and has established accepted methods for doing so, as Bennett points out in his discussion of a theory of action:

> The notion of theory of action is not often applied to assessment programs because such programs are not generally intended to cause change in individuals or institutions in the same sense as with an educational or social services intervention. However, in educational accountability testing, as, for example,

conducted under No Child Left Behind (NCLB), change *is* intended, at least at the school level where sanctions for poor performance are applied (Bennett, 2010, p. 71).

Certainly, current efforts to create comprehensive and inclusive assessment systems in the context of the RttT programs seek ambitious change for all students. AEE advocates comprehensive assessment systems embedded in the instructional milieu. In the literature, technical considerations regarding learning models refer to validating the progressions themselves (Corcoran, Mosher, & Rogat, 2009),—an issue that has additional implications when special populations are considered. For example, should the learning models present typical and alternative pathways? Will particular special populations (e.g., ELLs, students with learning disabilities) demonstrate unique learning pathways that can be differentiated from their "typical" peers' pathways? If so, what are the implications for making inferences at varying levels of assessment results?

Validity in the Context of Comprehensive Assessment Systems

In reference to assessments, "*Validity* refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations" (AERA, APA, & NCME, 1999). Test validation seeks not to validate the test itself but to validate interpretations of the test scores for particular purposes and uses. Cronbach (1988) and Kane (2006, 2008) suggest using an argument-based approach. The interpretive argument specifies the proposed interpretation and use of test scores. The validity argument provides a critical evaluation of the overall plausibility of the interpretive argument.

Aspects of Validity

Of concern are legitimating learning progressions, evaluating the reliability of comprehensive assessment systems, validating inferences based on assessment results, and using the results to guide improvement in outcomes for students in special populations. The three validity concerns regarding organized learning models and developing assessments based on them follow: (1) the legitimacy of the models themselves; (2) the validity of basing assessment system designs on the learning models; and (3) the validity of inferences from assessments designed and developed with learning models as the framework.

The three-dimensional rectangular box depicted in Figure 1 displays the nuanced concerns. The horizontal slices represent validity in relation to the organized learning models; the vertical slices, front to back, represent validity in relation to the type of assessment; and the vertical slices, side-to-side, represent the validity for particular student populations.

Evidence-Centered Design

Special populations can challenge assessment systems designed to measure achievement in the academic domains of English language arts, mathematics, and science; and concerns have been expressed about reliability and validity in measuring the academic achievement of special populations such as individuals with disabilities and ELLs (Hansen & Mislevy, 2008; Shafer,

2005; Tindal et al., 2003). The evidence-centered assessment design (ECD; Mislevy, Steinberg, & Almond, 2003) framework makes building assessment arguments explicit and provides tools for doing so (Mislevy & Riconscente, 2005). Assessment is conceived as an argument derived from imperfect evidence that aims to make the claims explicit (i.e., intended inferences based on scores) and the nature of the evidence that supports those claims (Hansen & Mislevy, 2008). In this complex validity environment, ECD has been shown to facilitate assessment design.

Figure 1. Aspects of validity in the context of comprehensive assessment systems, special populations, and learning models.
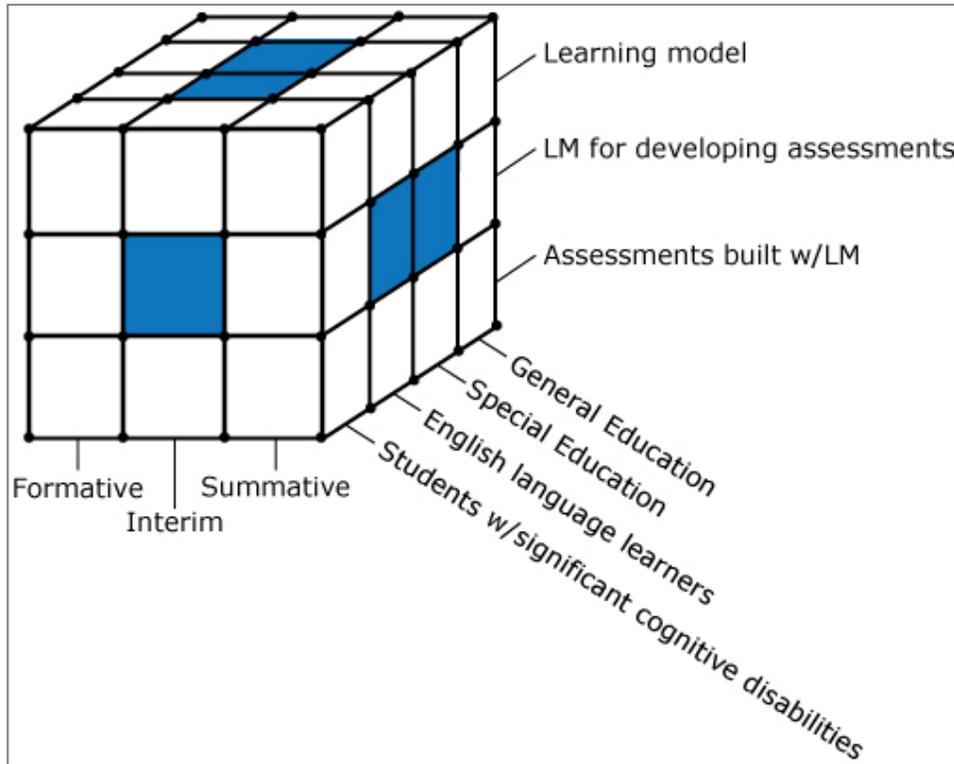
Zhang, Mislevy, Haertel, Javitz, Murray, and Gravel (2010) point out that the validity of inferences about what students in special populations know and can do entails a variety of inferences and assumptions that need to be made explicit in the assessment argument. The Zhang team began with a theory-based approach to developing quality assessments of scientific inquiry; it combined developments in cognitive psychology and research on science inquiry with advances in measurement theory and technology—Principled Assessment Designs for Inquiry (PADI). Using the PADI-Special Education (PADI-SE) program of research, they extended their work in ECD; they sought to explicate the chains of inference and their associated assumptions concerning special education students by applying evidentiary reasoning to address the complexities of the validity argument associated with accessibility features. They also sought to describe evidentiary structures that capture prominent aspects of the validity argument and to determine assumptions about students' capabilities and task requirements.

Using PADI-SE, they developed a design method for improving the validity of inferences about the performance on large-scale science assessments of students with disabilities. They combined the use of universal design for learning (UDL) with ECD to redesign statewide science items (Haertel et al., 2010) to evaluate the knowledge and skills of all students more accurately, including students with high-incidence disabilities, mild mental retardation, and learning disabilities.

Technical Documentation

Knowing What Students Know (National Research Council [NRC], 2001) uses Australia's Developmental Assessment as an example of "models of learning and *progress maps*, [that are] intended to serve as a basis for the designing of both large-scale and classroom assessments." According to the authors, "the Developmental Assessment approach represents a notable attempt to measure growth in competence and to convey the nature of student achievement in ways that can **benefit teaching and learning**[1]" ((NRC, 2001, p. 190).

OLM development follows an iterative process that begins with expert experience and research, followed by a review of research on learning in the domain, and then articulation of the sequence of development. Initially, teachers evaluate the sequence, with student responses to assessment tasks based on the sequences analyzed and discrepancies reported to developers so that they can refine the sequences. Thus, a process informs the sequences as evidence informs the learning model.

Providing guidance about "how to" develop OLMs is beyond the scope of this white paper; however, the recent literature describes processes for developing OLMs (i.e., learning progressions and LMs) (Alonzo & Steedle, 2009; Heritage, 2008, Hess, 2008). Processes for creating *inclusive* OLMs, or models that explicitly consider the acquisition of content by special populations have received considerably less attention. To develop an OLM, a series of decisions are made about the domain, the approach (e.g., top down, bottom up), the span of the progression, the level of detail or granularity (Heritage, 2008), and whether the assessment is a formative classroom one or a large-scale one. General steps adapted from CPRE Research Report # RR-63 (Corcoran, et al. 2009) for developing an OLM are sketched below:

- Decide to develop an OLM as the foundation for instruction and assessment
- Choose an approach to construct the OLM: progress maps, top down, bottom up, etc.
- Employ an iterative development process
- Refine the OLM on the basis of empirical evidence
- Develop the assessment based on the OLM
- Administer the assessment to validate the OLM.

---

[1] Emphasis added.

Technical considerations call for documentation (Downing & Haladyna, 2006) of: the design, development, and implementation of the assessment process employed; the considerations and decisions made and their outcomes; the theory used; the evidence gathered; and adjustments made over time. This iterative process for development resembles Design Based Research, a systematic but flexible method for improving educational practices through iterative analysis, design, development, and implementation that is based on collaboration among researchers and practitioners (Wang and Hannafin, 2005). . Concerns to consider are summarized below.

Considerations in selecting approaches, processes, and methods for developing OLMs for inclusive assessment systems

Learning Theory. The different learning theories used to develop OLMs may affect how the resulting learning progressions are expressed. Some may want to develop only one OLM that can be used for all student populations. Such a single OLM is based on the assumption that the OLM includes all possible paths or only the dominant paths. Others may want to develop different OLMs that are based on the same learning theory and content standards to ensure that the OLMs relate to all populations.

Alternate Pathways. Whether or not an OLM explicitly models alternative pathways, it is critical that its developers carefully consider learning theory and models that apply to special populations and thoroughly analyzing those populations' potential alternative routes to acquiring content. Consideration of access needs and other supports is fundamental to the development of meaningful pathways and will help ensure that the OLM is applicable to *all* students.

Accessing Diverse Expertise. Meaningful representation of all students' processes for acquiring content requires a multidisciplinary approach to construct OLMs. Because OLMs reflect the complex interaction among cognition, curriculum, and instruction, several types of expertise (especially during the initial stages of OLM development) are needed. Development teams may include special educators, assistive technology specialists, ELL teachers, curriculum specialists, general educators, and cognitive scientists, as well as speech-language pathologists, given the complex communication needs of special populations.

Flowers, Wakeman, Browder, and Karvonen (2007) provide a useful example for determining the expertise needed to ensure the needs of special populations. Their assessment alignment work specified that the alignment team must include at least two academic content experts, two experts in curriculum for students with significant cognitive disabilities, and one alignment leader for each subject area across all the grade levels.

Iterative Approach. Literature documenting how students learn particular concepts and content over time is scant (Alonzo & Steedle, 2009), especially so for populations with significant disabilities. In the absence of empirical research, OLM development necessitates an iterative process, in which several individuals reflect on models after they have been constructed. Popham (2011) proposes conducting a "backwards analysis" of developed learning progressions by moving back through proposed routes or building blocks to skill acquisition and checking to ensure that the building blocks are appropriate and in the correct order(s) of acquisition. Initial learning progressions should be considered conjectural (Alonzo & Steedle, 2009) and require additional research to validate their claims.

Alternate Assessment. When considering alternative assessment based on alternative achievement standards (AA-AAS), the content needs to be clearly articulated; at the same time,

the content may be "restricted in scope or complexity or take the form of introductory or pre-requisite skills" (U. S. Department of Energy, 2005, p. 26). For OLMs developed for students who have complex needs, multiple pathways may need to be articulated to accommodate their particular learning characteristics. Some students may need access points and supports embedded in the OLMs. Thus OLM development requires evaluation for legitimation.

Documentation of Technical Considerations

We have discussed two types of technical considerations related to OLMs: (1) validation that accounts for the particular nature of OLMs designed to describe the learning of special populations; and (2) validation of assessments based on the OLMs.

In both cases, developers need to ensure that OLMs and assessments meet applicable professional and technical standards. The *Standards for Educational and Psychological Testing* (1999) provides applicable requirements for validating assessments. Technical documentation of tests includes evidence of reliability and validity that supports inferences that will be made from assessment results. Technical qualities for many educational tests also include construct representation and curriculum relevance (Messick, 1989). Technical documentation of OLM legitimacy should:

- Indicate the processes and results throughout the development and validation process, with description of the procedures used to develop the learning models, the criteria applied in selecting those procedures, and the analysis conducted to address the legitimacy of the models

- Describe the relationship between the learning model and the results of the assessments based on the learning model, indicating whether assessment results confirm the model or whether revisions are needed to address the learning characteristics of certain special populations

- State which additional validation studies are needed to legitimate the progressions, whether longitudinal studies need to be conducted, and whether differences in the models favor one path for one subgroup and another model or path for another subgroup.

Validating OLMs

This section describes two approaches developed to legitimate (or validate) OLMs. The first approach by CPRE (2009) describes a systematic and iterative process that includes field testing, refining learning models based on evidence, and empirical studies designed to evaluate whether instruction based on the learning models led to improved mastery when compared with conventional instruction. The second approach is a work in progress, being developed and implemented by the Dynamic Learning Maps Alternate Assessment System Consortium, a 13-state consortium funded by a grant from the U.S. Department of Education to develop an alternate assessment system for students with significant cognitive disabilities. The project is completing its first phase—developing dynamic learning maps (DLMs).

Validating Learning Progressions in Science

The CPRE report describes how the OLMs were developed. They began with systematic examinations of relevant theory and research about how students learn the target concept. They asked questions such as:

- What do we know about the typical student's entering understandings, misunderstandings, and misperceptions?
- What do we know about the struggles students have with a particular concept?
- What have we learned about the steps students typically go through?

For measurement and special populations, technical considerations are sometimes modeled on those used in assessing achievement and learning in general education. One example, for which several approaches have been used (Rothmnan et al., 2002; Webb, 1997), is the study of the alignment of assessment content and items with the academic standards adopted by a board of education. In 2008, Flowers et al. developed procedures for evaluating the alignment of alternate assessments based on alternate achievement standards that included areas of study common to other studies of alignment; they also developed additional areas of study tailored to the learning characteristics particular to students with significant cognitive disabilities. They examined alignment methods that had been recognized by the field, as a place to start (Porter & Smithson, 2001; Webb, 1997; and (Rothman, Slattery, Vranek, & Resnick, 2002). Their alignment methods, called *Links for academic learning: An alignment protocol for alternate* that had been recognized by the field (Porter & Smithson, 2001; Rothman, Slattery, Vranek, & Resnick, 2002; Webb, 1997). This model of adapting proven methods and supplementing them with features tailored to the particular population provides an example for developing methods for legitimating OLMs. Theory and research findings guided development rather than conventional wisdom or consensus processes.

The CPRE (2009) report describes the validity claims for the assessment derived from its OLM, explaining that validating assessments based on OLMs is part of the process of developing those models. Construct validity requires that assessments report validly on students' levels of progress and that they discriminate reliably and validly among levels. Further, the assessments need to explicitly define how students at each level express their knowledge and skills. Assessments derived from s are also expected to positively influence instructional decisions.

From CRPE's perspective, learning progressions are legitimated by gathering evidence through testing, but they can also be discredited on the basis of evidence. Field tests are conducted to determine if, in fact, most students do follow the predicted pathways when they receive appropriate instruction. On the basis of field testing results, learning progressions are altered and refined. Finally, studies can also be conducted to determine if instruction based on the learning progressions is better than conventional instruction for improving mastery of learning targets. The report introduces a different perspective on validity from the one we have been applying in validating inferences based on assessment scores. CPRE assessments are derived from learning progressions and then employed to validate the progressions. The focus is on improving instruction and learning. Will OLM development produced with special populations in mind employ a similar approach? What will be the similarities and differences?

Developing and Validating DLMs for students with significant cognitive disabilities.

The Dynamic Learning Maps (consortium is completing the initial development of two DLMs that are designed to show a learning landscape in which multiple skills relate to many other skills, as well as multiple learning pathways. The iterative process for building the DLMs is not differentiated from the process for validating the maps and both are embedded in the development process. The sequence listed in Table 1 and discussed in the following text describes the steps from initial DLM development to mature DLMs. More formal steps in legitimating the DLMs in the development process are flagged in the table as validation steps. We note that an iterative process may move forward incrementally as evidence and information gathered along the way inform the process. Listed sequentially, the steps show the route toward validation.

Organizing Principles (1). A unique characteristic of OLMs developed in the context of next-generation comprehensive assessment systems funded by the U.S. Department of Education is their grounding in CCSS. The four consortia of states building next-generation assessment systems with grant support from RttA and GSEG programs are all developing assessments based on the CCSS. The first step in the DLM approach entails using the CCSS as a framework for identifying organizing principles that guide research literature searches.

Research Literature Review (2). In the second step, which can be conducted in tandem with the first step, research is reviewed about how learning regarding content within domains occurs, particularly in reference to targeted knowledge and skills. Two areas are captured: how skills are developed and precursor skills that lead to school-age academic skills. Understandings obtained from this review are applied to development and used to inform the nodes in the DLMs. This aspect continues throughout DLM development and as empirical data are accumulated.

Table 1. Steps in Developing DLMs

Steps in Developing DLMs	Validation
1. Identify Organizing Principles	
2. Review the research literature	
3. Develop the DLM nodes, placing nodes and edges (pathways)	
4. Compare the DLM with CCSS	
5. Discuss and evaluate sections of the DLM in internal meetings	
6. Conduct an external review of the DLM with curriculum experts/master teachers in the domains	Yes
7. Conduct an external review of the DLM, evaluating alternate pathways and completeness of nodes	Yes
8. Review the DLM for concordance within learning maps, essential elements, and instructional achievement level descriptors	
9. Conduct cognitive/observational labs	Yes
10. Review by research experts	Yes
11. Use evidence-centered design to support node differentiation	Yes
12. Conduct a small sample pilot test to determine if assessment data support (1) the differentiability of select nodes, and (2) the hypothesized select pathways	Yes
13. Moderate the sample field testing to determine if assessment data support (1) the differentiability of all nodes, and (2) all hypothesized pathways	Yes

Develop and Assemble the DLM, and Position Nodes and Edges (3). Nodes, or skill-based targets, are developed based upon the results of the review of research literature, which has been structured around the organizing principles from the CCSS. Assembling the DLM involves positioning the nodes and establishing relationships between them by articulating pathways along them. The resulting DLM is a complicated network of nodes and pathways. The nodes differ in that some describe knowledge, some skills, and some concepts; it is also assumed that more than one path may lead to those different nodes. The placement of nodes and pathways documents information obtained from the literature or the best hypothesis about how learning occurs.

Compare the DLM to CCSS (4*).* Periodically during DLM development, content experts compare the map's nodes and pathways with sequences implied by the CCSS. Differences between the maps and CCSS-proposed sequences are noted and evaluated. Some differences may trigger re-examination of the research literature to obtain evidence for alternative pathways.

Evaluate DLM sections in internal meetings (5). Throughout the development process, evaluation sessions are held periodically to review DLM sections for concordance. Multiple, independent sources of evidence that are concordant with each other are expected to be mutually supportive and lead to much stronger conclusions than those depending on single sources. Verification of the links, sequences, and positioning of the nodes entails ongoing, continuous improvement.

Conduct an External Content Review (6). English language arts, mathematics experts, and master teachers review DLM content for its adherence to targeted knowledge, skills, and abilities. Facilitators conduct training, evaluate the reviewers for consistency of their ratings, and establish fidelity with the rating procedures. When disagreements occur, facilitators implement a systematic procedure for reconciling differences. Recommendations regarding revisions or re-examination of content are documented and returned to the DLM developers for resolution. Experts in English language arts, mathematics, and cognitive scientists also closely examine the precursor skills that make up the foundation of the map. The research team examines the reviewer feedback and makes determinations about how to strengthen the map.

Conduct an external positioning review (7). A systematic consensus process is the principal method for resolving concerns regarding positioning during DLM development. Before data collection, a series of ongoing reviews by qualified professionals is employed to ensure that development adheres to the principles established by the research design. Ultimately, evidence based on simulations, pilot investigations, and field testing provide data to support the continuous improvement cycle intended to ensure the validity and legitimacy of the DLMs for students with significant cognitive disabilities.

Review for Concordance (8). English language arts, mathematics experts, and master teachers review the DLMs for concordance, essential elements, and descriptors of instructional achievement level. Again, facilitators conduct training, evaluate reviewers for consistency of ratings, and establish rater fidelity with the rating procedures. The desired outcome is that the multiple components are mutually consistent.

Conduct cognitive/observational labs (9). Tasks are developed on the basis of the DLMs. Task tryouts are conducted with a purposively chosen sample of students to verify the credibility of the nodes and pathways. Students are sampled to represent a variety of learning characteristics, levels of intellectual functioning, and formats for communication. On the basis of

these tryouts, nodes and pathways are revised and annotated before a formal research expert review takes place.

Review by Research Experts (10). This third-party, independent review is conducted by a team of experts who are external to the development process. The reviewers are chosen to represent expertise in the following areas: English language arts, mathematics, special education, cognitive psychology, measurement (particularly classroom measurement), and test development. The external review findings and recommendations are presented to the DLM Technical Advisory Committee (TAC) for review and comment. Once TAC members weigh-in the DLM team decides on changes and makes adjustments to the maps.

Implement the Evidence-centered Design (11). The DLM adopts a design method for improving the validity of inferences about the performance of students with significant cognitive disabilities by combining the use of UDL with ECD to design tasks for learning targets that are accessible to those students. In this part of the validation process, nodes, pathways, and tasks are examined in the ECD context to evaluate the ability of the alternative nodes and pathways to meet the varied levels, communication pathways, and foundational or precursor skills required for the targeted knowledge, skills, and abilities.

Conduct Small Sample Pilot Testing (12). An important question regarding the nodes and pathways in DLMs is, "Do the nodes and pathways represent likely pathways that students with significant cognitive disabilities would navigate on the path to learning targeted academic content." To address this question, pilot testing with samples of 20-50 students per studied node is conducted to determine if the pilot data support the differentiability of those nodes and the hypothesized pathways associated with them.

Moderate the Sample Field Testing (13). Using the results of pilot testing, field tests with samples of 100-500 students per node are developed to determine the degree to which data support differentiability of nodes and hypothesized pathways.

Reflections on Legitimating OLMs

The CPRE approach for learning progressions in science and the approach for legitimating DLMs for students with significant cognitive disabilities differ significantly. The science example entails a complete cycle: it begins with development of the progression, followed by verification of the progression by administering assessments based on that progression to determine whether the resulting instruction improves instruction and learning. The CPRE report describes two validation approaches—a learning progression for the carbon cycle involving organic compounds and a learning progression for modern genetics.

The DLM project has taken on a formidable task. The DLMs are founded on CCSS across the grades and are informed by research and expert counsel. The DLMs are intended to align with CCSS and to provide extensions that serve students with significant cognitive disabilities. The multistate consortium is building the initial DLMs, and current validation efforts consist of quality assurance procedures embedded in the development process. The DLM project's scope is much broader than that of the CCSS and more varied, including typical pathways and alternative pathways.

These two approaches do not cover all the different approaches researchers are using to validate OLMs, but they do shed light on the range of approaches under way.

14

Considering the Evaluation Method as an Approach to Validating OLMs for Special Populations

The evaluation method is similar to the logic models derived from theory-driven evaluation processes (Chen & Rossi, 1983; House, 1975; Kellogg Foundation, 2004). Theory-driven evaluation begins with defining the theory and then describing how programs work and what the interim and long-term outcomes and goals are. These models are often used when statistical analyses of outcomes are difficult to obtain.

The number of students with similar complex needs in special populations is not large, and few statistical measures lend themselves to validating OLMs for those populations. Thus, validating OLMs might use of an argument approach.

Kane's method for constructing a validity argument, propositions, claims, and evidence (2006, 2008) may provide an approach. For example, propositions might be developed if:

- The OLM represents a learning theory that is supported by available research
- The grade/content standards on which the OLMs are based define the learning expectations
- The nodes are organized to represent progressions and relationships among the content for defined populations
- The learning supports and access points needed by some students in special populations are used correctly.

If these conditions are fulfilled:

> The OLM may be appropriately used to develop instructional and assessment modules and tests for the complete population of students, including special populations.

Examples of possible claims and evidence for these propositions follow:

- **Proposition 1.** The OLMs are based on a learning theory. *Design Claims*: clarity, feasibility, assumptions of progressions designed. *Evidence*: literature review of learning theory and how students learn, expert review.
- **Proposition 2.** The OLMs are based on grade/content standards that define learning expectations. *Design Claims*: clarity, alignment of the OLMs with standards. *Evidence*: expert review, alignment studies.
- **Proposition 3**. The nodes are organized to represent progressions and relationships among the content standards for a defined population. *Design Claims*: clarity, alignment of OLMs with standards. *Evidence*: literature review; expert review; alignment studies, if data are available; calculation of statistical relationships such as correlations or probabilities among nodes; if more than one OLM is based on the same theory and if content is developed for different populations, review for consistency.
- **Proposition 4.** Learning supports and access to content are defined as necessary to assist users to understand and use the OLMs. *Design Claims*: clarity, accessibility (so that students have consistent access to communication systems that allow them to respond to academic information). *Evidence*: expert review, research studies on accessibility and use.

- **Proposition 5.** The OLMs inform the development of instructional modules and assessments. *Design Claims*: clarity and alignment of OLMs with material as evidenced by a model such as *Links for Academic Learning model* (Flowers & Browder, 2010), and assessments. Evidence: expert review, alignment studies relating OLMs, instructional modules and assessments, research studies on use.

Including supports in the OLMs. Academic content standards are educational targets that outline what students are expected to learn at each grade level. Teachers ensure that students work toward grade-level content standards by using a range of instructional strategies based on the varied strengths and needs of students. For students with disabilities, accommodations are provided during instruction to help promote equal access to grade-level content.

Common supports for students with significant cognitive disabilities are content scaffolding, augmentative and alternative communication devices, task analyses, picture symbols, and prioritization of learning goals. The supports help students access the data and enter into the OLMs. Not all supports are required for all students. Supports, as with any accommodation, should be specific to students' learning styles and needs. The supports that instructors use assist students in showing what they know, but supports should not introduce construct irrelevant error, such as that which results from excessive prompting.

Access points in the OLM content need to be defined in the OLMs. Like supports, they are not necessary or required for every student for each node. They should be determined by the students' teachers and other instructional support staff, such as special educators. The OLMs need to be developed with accessibly considered. If access points are not included, some students will not be able to enter into the OLM to show what they know and can do.

Summary. OLMs developed to meet the instructional and assessment needs for students in special populations may require adapting methods for evaluating validity. Some constraints are related to: (1) the limited size of subpopulations, which might limit the types of analyses available; (2) the diversity of the subpopulations, requiring supports to be personalized to individuals; and (3) the relationship between the targeted content and the supports, which may alter the intended assessment target. Other constraints, apart from these examples, in developing OLMs with special populations also apply.

Building Inclusive Assessments based on OLMs

The New Measurement Paradigms (NMP) group is exploring a new concept for educational measurement (CADRE, 2012). The group points out that computer-based learning environments create new demands for measurement methods by increasing the ways in which a student can respond during an assessment task and by requiring that assessments be conducted "on-the-fly" so that assistance can be provided to learners as they progress. The group speculates that the old measurement models, which typically worked with full sets of response data at the end of a set of tasks, are becoming outmoded. NMP learning models are being developed and tested in intelligent learning environments that seek to carefully measure and promote academic growth in students. This approach emphasizes judgments about students' levels of knowledge and skills while they are learning, rather than judging achievement based on summative "passing" scores.

These test development concepts depart from the 12 steps for effective test development (Downing, 2006) and represent innovative approaches. Designing and developing assessments founded on OLMs may differ from "traditional" test development, which is related to and often concurrent in the process for validating the OLMs themselves. This section considers the implications for building assessments based on OLMs.

Purposes and Types of Assessments

Assessments developed to show performance along a progression from novice to master can indicate how instruction could be changed on the basis of the assessment results (see Table 2). Perie, Marion, Gong, and Wurtzel (2007) relate the technical considerations for assessments to their purposes (also described by Bechard et al., 2012):

- Formative assessments intended to inform instruction relatively immediately
- Interim or benchmark assessments intended to inform instructional planning—a more group-oriented approach
- Summative assessments employed to inform curriculum planning, program evaluation, and accountability (usually at the school and district levels).

Wilson (2009) distinguishing between formative and summative assessments indicated that one major difference between the two was that formative assessments seek to help learning by providing feedback to teachers and students for use in modifying teaching and learning, whereas summative assessments provide a summary of what a student knows, understands and can do, rather than providing feedback to modify the teaching and learning. He asserts that formative assessments can assume many curricular grain sizes, ranging from assessments that are narrowly diagnostic to those that are broader and provide "curricularly diagnostic" information to a teacher.

Thus the level and grain size of the learning model the assessment focuses on are based in part on the purpose of the assessment. Different types of assessments may be appropriate for different parts, levels, or grain sizes. The knowledge desired about students from the assessment governs its design. Although this is true for any assessment, with OLM-based assessments the knowledge desired about students may differ from that provided by traditionally developed assessments. This is almost a tautology – OLMs are based on the assumption that the knowledge desired about students differs from that which assessments now provide.

Designing Assessment Items and Scores to Reflect the OLM and the Purpose of the Assessment

Design and development of a comprehensive assessment system that is inclusive of special populations based on OLMs need to address all the diverse learning characteristics of the population of test takers (see Table 2).

Table 2. Distinctions between formative and summative assessment in the OLM context

Factors to consider	Formative Assessment	Summative Assessment
Key concepts	Constant feedback and information about students' progress, plus support for modifying instruction	Learning at one point in time to evaluate students' performance against a standard
Scoring models and OLM implication	Given formative assessment feedback for use in evaluating student progress and informing instruction results, more qualitative, indicating the degree of learning, strengths, and areas of weakness or gaps in understanding	Broadly sampled content to represent achievement typically employing multiple-choice, short-answer, and constructed responses that can be scored correct or incorrect; a score that can be interpreted as a level of achievement or mastery
Consideration for this type of assessment related to OLMs	More fine-grained, designed to reveal misconceptions and incomplete understandings; narrowly focused and diagnostic	Comprehensive, covering a course completed or a grade level's worth of content—sufficient coverage to support reliable (reproducible results)

In designing the assessment system, perspective, inclusive ECD features should be made explicit in three areas:

- The Presentation Process in which the task is presented and the response is captured
- The Task Models in which the situation, presentation material, work products, and task model variables are specified
- The Item in which the particular presentation and response features for a particular item are designated.

The following development areas need to be incorporated into the assessment design from the outset.

Access Options and Tools. When items or tasks are presented to students, the stimulus and options need to be available in alternative presentation formats, (e.g., text, audio [text-to-speech]); an image or description may be attached to an item in computer-administered assessment tags to indicate presentation options. The presentation process, for computer-administered assessments, may be delivered via a computer screen and speakers, and the student's response may be captured through a key stroke or a voice capture. Such options need to be made explicit in the design and development of the presentation process, as well as in the task model, with items coded appropriately in the item database.

Scaffolding and Support Options. These options are particularly useful in a formative assessment grounded on a developmental sequence of points along a progression. Performance on the earlier (lower level) points in the sequence might indicate partial or preliminary understanding. When tasks and items are administered with varying levels of support, information about partial knowledge or emerging understanding needs to be coded.

Assumptions about all Types of Students, Particularly Students with Disabilities and ELLs. At multiple steps and phases in the assessment development process, including guidelines

for item writers and descriptions of processes, assumptions that the population test takers include students in special populations need to be made explicit.

Universal Design Principles within Appropriate Processes. Three guidelines ground the universal design principles: present information and content in different ways, differentiate the ways that students can express what they know, and stimulate interest and motivation for learning. These guidelines should be adopted at the outset of the assessment development process.

Assurance that Access Tools and Supports Are Appropriate to the Area(s) along the LP Being Measured. This assurance requires considering the effects of the access tools and supports for the construct being assessed. In some cases, an access tool (e.g., an audio text-to-speech presentation) may "give-away" the correct answer by reading aloud numerals or symbols that are part of the construct being assessed.

Identification of Entry Points. When starting with a developmental sequence of learning provided by OLMs the point to begin assessing needs to be guided by the assessment purpose. Determining the level to inform instructional planning may include establishing both what a student knows and where understanding is just beginning to emerge.

When assessments are founded on OLMs, assessment developers may have a better understanding of the nature of and performance from different OLM entry points into assessment tasks. Considering the OLM and the unique characteristics of students in special populations, developers need to consider how tasks and items can be made appropriate for and accessible to promoting learning for students with disabilities and ELLs. The assessment should be designed to allow both for alternative entry points and for multiple pathways.

Technical Considerations of OLMs in Accountability Systems

OLMs containing student outcomes that are used in instruction and assessment primarily serve to help assure that all students learn essential content standards. Thus, the OLMs guide the instruction and assessment for students, whether that assessment is formative (suggesting modifications in instruction and student learning) or summative (gauging the nature and extent of student learning of these concepts).

However, summative assessments can also be used to hold educators and/or schools accountable for student learning. Because schools are funded to help students learn, policymakers believe that it is fair to use the assessment results to hold schools accountable for student achievement. Accountability remains an integral aspect of assessment even as we move toward comprehensive and inclusive next-generation assessment systems. Therefore, we review recent accountability determinations here.

Models for School Accountability. States and districts hold schools accountable in two basic ways:

The *status* of student achievement. These accountability systems define a level of adequate student achievement. Then, the desired levels of overall student performance at the school, district, and/or state level are also defined. These definitions are sometimes numerical only, or they may entail written definitions called achievement-level standards or performance-level

standards. Note that these are generalized narrative descriptions of what students need to know and can do to score at each achievement or performance levels.

Typically, the overall performance of all students is summarized and used to determine whether or not schools meet or exceed the desired targets for schools in the state. The convention has been for states to classify schools on the basis of the percentage of students passing the state's summative assessment. Schools that exceed the target(s) are considered to be "successful," whereas those that do not are viewed as in need of improvement. Note that the labels applied (and their definitions) vary from state to state.

Growth in student achievement is another way for holding schools accountable for the *changes* (improvements) in student achievement that occur over time, such as from third to fourth grade. Rather than expecting all schools to meet or exceed a specified level of performance (i.e., a defined status level), schools can also be expected to help all students gain a specified amount of learning (defined by gains on the assessment), regardless of the level of initial performance on the assessment.

Issues in Each Accountability Model. Each accountability model entails challenges and provides benefits in holding schools and educators accountable, as described below:

- **Status Model Benefits and Challenges.** The major advantage of the status model is that it answers the basic question that often drives policy makers to implement accountability systems: Are all students scoring at a satisfactory level or above? However, the challenge in using this sort of metric for accountability is that schools that do poorly might feel it is impossible to improve and thus cease trying therefore depriving students of needed assistance.

- **Growth Model Benefits and Challenges.** The benefit of using growth or changes in student performance for accountability is that it serves to encourage educators to help students at all levels of achievement to improve their performance. This type of accountability system could help assure that educators assist both low- and high-achieving students. A negative aspect of this type of system is that educators may be rewarded for students who have gained substantially, but whose current level of performance is inadequate. Thus, parents and policy makers may be critical of educators who appear satisfied with student gains to levels that are still unsatisfactory.

- **A Potential Hybrid Accountability System.** Although discussion about the nature of accountability systems often focuses on which of the two accountability models described above is preferable (and the rationale for its selection), it may be more useful to combine the two approaches, thus encouraging educators and schools to strive for high levels of student proficiency (through a "status" measure) as well as seeking performance improvement for all students (i.e., a "progress" measure). Thus, the use of the two measures together may permit the following sorts of statements to be made about a school:

School 1: Most students (87%) meet or exceed the state's target for proficiency, but improvement has been modest over the past 3 years.

OR

School 2: Most students (87%) meet or exceed the state's target for proficiency, but student performance has substantially declined over the past 3 years.

School 3: Few students (24%) scored at the proficient or above level, but improvement in student performance over the past 3 years has been significant.

OR

School 4: Few students (24%) scored at the proficient or above level, and student performance has shown little improvement over the past 3 years.

As can be seen, Schools 1 and 2 and Schools 3 and 4 are very different, even though each set of schools has the same status scores. Schools 1 and 3 both show improvement in performance, and Schools 2 and 4 do not. Schools 1 and 3 differ, however, due in large part to their differences in proficiency. Therefore, using both a status and progress model for accountability makes these important differences evident.

Issues in Using Assessments for Both Instructional Improvement and Accountability Purposes. For several reasons, assessment programs and the results they produce might be used both to improve student learning, and to hold the schools and educators accountable for student performance. This appears, on the surface, to be a logical way to reduce the use of redundant assessment programs, as well as to extend the use of the achievement results. However, there are several cautions in doing so.

First, when educators are held accountable for the achievement of students, they may redirect their instruction focus only to those skills that are assessed, neglecting other important skills that students need to be taught. Moreover, attention to entire content areas that are not assessed (e.g., the arts, social studies health education) may be absent.

Second, pressures to have students do well on the accountability assessments may distort teachers' instructional methods and assessments. Rather than providing learning experiences that are student-centered, teachers may feel pressured to narrowly focus on specific content and skill development, through lectures and drill and practice worksheets. Doing so may result in short-term learning, but fail to instill deeper understandings in the content area (and, perhaps, lead to later learning problems because the lack of conceptual understanding can lead to inability to learn more complex skills).

Third, the pressures on educators may translate to attempts to "game the system" by having special population students not participate in the regular assessment (because they may not do well on these assessments). This may result in students being assigned to AA-AAS, even though this assessment is not suitable for them academically. Occasionally, educations may attempt to exclude the students from the accountability assessments altogether.

Fourth, the pressures on educators may translate to pressures on students, who are not permitted to explore new ideas or relationships between new material and their prior knowledge because doing so may take too much time, given the many standards that need to be covered and the limited time to do so.

21

Our discussion of OLMs may chart a route to the future. OLMs indicate pathways students take toward achievement in a subject area, and they could be used in developing an assessment system that measures student progress toward college and career readiness. They could also provide a way to measure growth, by showing how students move forward along the progressions.

Therefore, it may be important to view the accountability system and its assessments as separate from the assessments designed to improve instruction. To be valid, learning progressions must be based on evidence. At this point, research referenced in Topic 1indicates progressions in learning in some areas, but not all. Additional research would be vital to support the development of assessments that yield valid information on student progress toward college and career readiness. Ideally, teachers would gather classroom-based formative and summative information that guided and evaluated student learning.

Unanswered Research Questions

Although OLMs hold great promise as a basis for assessment and accountability, many research questions still need attention. These questions fall into at least seven major areas: (1) the OLM development process, (2) OLM features, (3) OLM structure, (4) OLM invariance properties, (5) statistical methods for validating the OLM, (6) OLM use in assessment, and (7) measuring growth with an OLM. Examples of research questions follow:

The OLM Development Process

1. Does it matter whether the OLM is built from top to bottom, bottom to top, or from the middle out?
2. What skill sets and experiences are required for OLM review?
3. Should research or a policy serve as the basis for OLM development (like the CCSS)?
4. Should reviewers consider the OLM as a whole or look at and stitch together various parts of the model?
5. Would a research variant of crowdsourcing work to build out an OLM?

OLM Features

6. What grain size is most appropriate for using an OLM in assessment?
7. Should the single most common pathway or all pathways be mapped?

Structure of the OLM

8. Is the structure of knowledge best captured as many small linear progressions, a large acyclic graph, or a graph that allows cyclic subparts
9. Are aspects of structure like mathematics facts or vocabulary best accounted for by a traditional OLM or by grafting continuous variables into a hybrid model?
10. Do large OLMs need to be divided into neighborhoods for estimation or reporting?
11. Is it better to define all nodes to be accessible or to have alternative pathways to address the different ways students access information?
12. What node size communicates most usefully for teachers?
13. Is modeling all paths necessary (i.e., without one, mastering a successor is impossible), or should compensatory nodes be modeled too?

14. Is there a maximum number of pathways that should come into a node or flow out of a node?

OLM Invariance Properties

15. Are the probabilities of mastering a node conditioned on mastery of surrounding nodes similar for all subpopulations?
16. To what extent do particular methods of instruction affect the conditional probabilities of mastery?
17. Are dominant pathways used by all students who have access to that pathway, or are many pathways relatively common sub-populations?

Statistical Methods of Validating the OLM

18. Given a set of data in response to questions that are connected to OLM nodes, is there a statistic that best captures the relative fit of the hypothesized set of connectors?
19. What statistical criteria can best help determine whether a particular OLM node is worth including?
20. What are the possible external criteria for validating a particular OLM?

LM Use in Assessment

21. How can an OLM best be used in the test development process?
22. Do OLMs add to the practice of evidence-centered design and thus improve the validity of inferences made from test scores?
23. Should teachers be presented with detailed OLMs, regardless of the OLMs' size and complexity?
24. Should students be presented with the actual OLMs, perhaps to serve as advanced organizers?
25. Should mastery of nodes (steps) in an OLM be presented as a score (e.g., percent of the OLM mastered, percent of salient neighborhoods covered, attainment of mastery of key nodes); if so, how should any inherent multidimensionality be captured?
26. What, if any, are the unintended consequences of the implementation of a particular OLM?

OLM Use for Accountability

27. Should the OLM's rich information be reduced to be consistent with other components of the accountability system?

Conclusion and Implications

Although the concepts embedded in OLMs have been discussed in an assessment context for more than 30 years, few, if any, large-scale assessments have systematically used such models. In this series of white papers, the authors have argued that OLM use holds much promise for improving the validity and utility of assessment programs. In this white paper we have explored some of the many technical issues in developing and using OLMs. Moreover we have highlighted the field's state of ignorance regarding many issues surrounding OLMs and presented a partial list of topics that should be researched.

Although several major CCSS assessment consortia are using OLMs, the assessment and accountability fields are littered with the unintended negative consequences of previous reform efforts. Measurement is approached with caution, and 20 years or more often passes from the development of an innovation to its operational use. With children's education at stake we do not want to make any mistakes. However, we must never forget that continuing to use imperfect systems also has negative consequences. Consortia are using OLMs. It is imperative that they document the technical considerations concerning their development and implementation.

References

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education, 93*, 389-421. doi: 10.1002/sce.20303

Alliance for Excellent Education. (2010, February). *Principles for a comprehensive assessment system* (Policy Brief). Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement, 8*(2-3) 70-91. doi: 10.1080/15366367.2010.508686

Bechard, S., Hess, K., Camacho, C., Russell, M., & Thomas, K. (2012).). *Why should cognitive learning models be used as the foundation for designing next generation assessment systems?* Symposium 2011 White Paper No. 1.1. Menlo Park, CA: SRI International.

Blank, R. K., Porter, A. C., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics & science*. Report from Survey of Enacted Curriculum Project (NSF REC98-03080). Washington, DC: Council of Chief State School Officers.

Chen, H. T. & Rossi, P.T. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*(3) 283-302. doi: 10.1177/0193841X8300700301

Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (CPRE Research Report # RR-63). New York, NY: Teachers College - Columbia University, Center on Continuous Instructional Improvement.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning.* Washington, DC: Council for Chief State School Officers.

Downing, S. M. & Haladyna, T. M. (2006). *Handbook of test development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Elementary and Secondary Education Act of 1965, Pub. L. No 89-10, *amended by* Pub. L. No. 103-382.

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards.* Charlotte, NC: National Alternate Assessment Center.

Haertel, G., Haydel DeBarger, A., Cheng, B., Blackorby, J., Javitz, H., Ructtinger, L., … Hansen, E. G. (2010). *Using evidence-centered design and universal design for learning to design science assessment tasks for students with disabilities* (Assessment for Students with Disabilities Technical Report 1). Menlo Park, CA: SRI International.

Hansen, E. G., & Mislevy, R. J. (2008). *Design patterns for improving accessibility for test takers with disabilities* (Report No. RR-08-49). Princeton, NJ: Educational Testing Service.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: Council of Chief State School Officers.

Hess, K. (2008). *Developing and using learning progressions as a schema for measuring progress*. Dover, NH: National Center for Assessment.

House, E. R. (1975). Accountability in the USA. *Cambridge Journal of Education, 5,* 71-78. doi: 10.1080/0305764750050202

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: Praeger.

Kane, M. T. (2008). Terminology, emphasis, and utility in validation. *Educational Researcher, 37*, 76-82. doi: 10.3102/0013189X08315390

Kellogg Foundation. (2004). *W. K. Kellogg evaluation handbook*. Battlecreek, MI: Author. Retrieved from http://opas.ous.edu/Committees/Resources/Publications/WKKF_EvaluationHandbook.pdf

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education/Macmillan.

Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology* (Principled Assessment Designs for Inquiry Technical Report 9). Menlo Park, CA: SRI International.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3-62. doi: 10.1207/S15366359MEA0101_02

Pelligrino, J., Chudowsky, N., & Glaser,R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system.* Washington, DC: The Aspen Institute.

Popham, J. (2011, April). *How to build learning progressions: Keep it simple, Simon!* Paper presented at the meeting of the American Educational Research Association, New Orleans.

Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing* (CSE Technical Report 566). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Schafer, W. D. (2005). Technical documentation for alternate assessment. *Practical Assessment, Research and Evaluation, 10*(10), 1-9.

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments* (Technical Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Timms, M., Clements, D. H., Gobert, J., Ketelhut, D. J., Lester, J., Reese, D. D., & Wiebe, E. (2012). *New measurement paradigms.* Newton, MA: Community for Advancing Discovery Research in Education.

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children, 69*, 481-494.

United States Department of Education, (2005). *Alternate achievement standards for students with the most significant cognitive disabilities: Non-regulatory guidance.* Washington DC: Author.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5-23. doi: 10.1007/BF02504682

Webb, N. L. (1997). *Criteria for alignment of expectation and assessments in mathematics and science education* (NISE research monograph no. 6). Washington, DC: Council of Chief State School Officers.

Wilson, M. (2009, December). *Assessment for learning and for accountability.* Paper presented at the ETS Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, Princeton, NJ.

Zhang, T., Mislevy, R., Haertel, G., Javitz, H., Murray, E., & Gravel, J. (2010). *A Design Pattern for a Spelling Assessment for Students with Disabilities* (Assessment for Students with Disabilities Technical Report 2). Menlo Park, CA: SRI International.