

Enhancing State Assessment Validity for English Language Learners with Disabilities

Kristin Kline Liu, Linda Goldstone, Martha Thurlow, Laurene Christensen, and Jenna Ward

National Center on Educational Outcomes – University of Minnesota

Enhancing State Assessment Validity for English Language Learners with Disabilities

Abstract

This paper compares and contrasts a set of expert-developed principles for including English language learners (ELLs) with disabilities in statewide content assessments and current assessment participation and accommodations practices reported by school and district staff in a consortium of five collaborating states (Minnesota, Michigan, Maine, Arizona, and Washington). The five states included have varying populations of ELLs in terms of size, percentage of ELLs with disabilities, and predominant home languages spoken by students. Still, the majority of the states' total ELL populations, and particularly ELLs with disabilities, are impacted by high levels of poverty and many of these students attend high poverty schools. By cross-walking "ideal" practices proposed by experts, and actual practices from practitioner reports, we highlight areas where local and state education agencies might implement changes to strengthen the validity of state assessments for this unique group of learners.

Enhancing State Assessment Validity for English Language Learners with Disabilities

The inclusion of students with disabilities (Spicuzza, Erickson, Thurlow, Liu, & Ruhland, 1996a; Thurlow, Ysseldyke, & Silverstein, 1995) and English language learners (August & Hakuta, 1997; Koenig, 2002; Kopriva, 2000; Spicuzza, Erickson, Thurlow, Liu, & Ruhland, 1996b) in large-scale assessment systems became a concern in the mid- to late 1990s. Some of these students had not been included in large-scale assessments before and were sometimes targeted for exclusion. A consequence of exclusion was that these students tended to not receive needed instruction because they were not participating in state assessments.

Only within the past decade or so has the importance of this issue been recognized for those students who are learning English and who also have an identified disability, referred to here as ELLs with disabilities (Thurlow & Liu, 2001). Although they do not represent a large portion of the K-12 population overall, ELLs with disabilities are a growing portion of the K-12 school population in almost every state (see www.ideadata.org, Tables 2-3, 2-3, 2-5a, 2-6a). With the increasing numbers of these students across the nation, addressing their needs, and ensuring that the approaches used to include them in large-scale assessment and accountability systems, is critical.

Title I and Title III legislation require that ELLs, including those with disabilities, be taught the same challenging content standards as their fluent-English speaking peers. Results from state-level content assessments show that ELLs with disabilities are among the lowest achieving students (cf. Liu, Barrera, Thurlow, Guven, & Shyyan, 2005; Liu, Thurlow, Barrera, Guven, & Shyyan, 2005). It seems straightforward to attribute poor student outcomes to the learning difficulties that ELLs with disabilities face. Yet evidence from educators, schools, and districts (Zehler, Fleischman, Hopstock, Stephenson, Pendzick, & Sapru, 2003) suggests that

instruction for these students is less closely aligned to state standards than instruction for fluent-English speaking students with disabilities or ELLs without disabilities.

State departments of education are making limited efforts to appropriately include ELLs and students with disabilities in statewide accountability tests (Albus & Thurlow, 2007; Altman, Lazarus, Thurlow, Quenemoen, Kearns, Quenemoen, & Thurlow, 2010), but still struggle with accurate identification of these students, accessibility of the tests, appropriate participation and accommodations decision making and appropriate score interpretation for ELLs with disabilities. These difficulties affect the validity of the assessment results when all students are included in the test-taking population, and the option of removing students who do not “fit” to improve the validity issue is no longer a viable option. The Improving the Validity of Assessment Results for English Language Learners with Disabilities (IVARED) project, a collaborative of five state departments of education led by Minnesota (Arizona, Maine, Michigan, Minnesota, Washington) and the National Center on Educational Outcomes chose to work together on a federally funded project¹ to specifically address the challenges of assessment validity for ELLs with disabilities.

The states that are part of the IVARED collaborative have already developed state content assessments in reading, mathematics, and science, and a few have already implemented Common Core state standards (and will implement future Common Core assessments) so the emphasis of our project is on maximizing assessment validity by improving testing decisions and practices, as well as the applicability of testing policies, rather than on developing new tests.

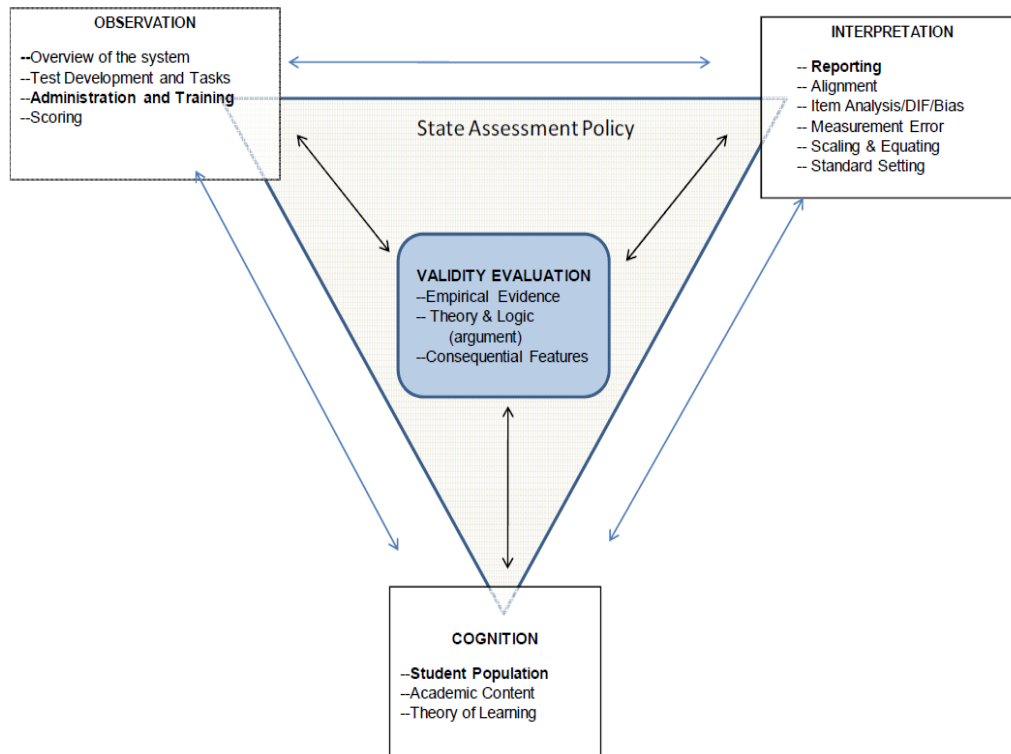
Conceptual Framework

¹ The Improving the Validity of Assessment Results for English Language Learners with Disabilities (IVARED) project is supported by a contract (State of Minnesota Award # B54419) to the National Center on Educational Outcomes. It is based on a grant from the Office of Elementary and Secondary Education, U.S. Department of Education (Award # EAG S368A100011). The grant was awarded to a consortium of five state departments of education (AZ, ME, MI, MN, WA) led by the Minnesota Department of Education. Opinions expressed in this paper do not necessarily reflect those of either the U.S. Department of Education or the Minnesota Department of Education.

What does it mean for an assessment to be valid for a particular population of students? According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), “validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9).

In keeping with the AERA definition of validity, Figure 1 shows how we conceptualized assessment validity as we began this three-year Enhanced Assessment grant project awarded by the Office of Elementary and Secondary Education. The conceptual framework on which the entire IVARED project, including the specific research studies described here, is an adaptation of Marion, Quenemoen, and Kearns’ (2006) version of the National Research Council (2001) assessment triangle. The triangle illustrates the components underlying appropriate reasoning about students’ knowledge and skills from test scores. Each of the vertices of the triangle contains one of the three interrelated components of reasoning based on evidence: Student cognition, observation, and interpretation. Marion et al. (2006) expanded the basic triangle to include the particular elements that states must consider within each of these three components as they develop, implement, and refine state assessments. Assessment validity is in the center of the triangle to emphasize that all of the vertices influence the validity argument for assessments.

Figure 1. Modified Assessment Validity Triangle



In Figure 1 student cognition is the bottom vertex of the triangle that supports the other two vertices. If educators and test developers are to make valid inferences about the learning of ELLs with disabilities using large-scale assessments they must have a clear grasp of the characteristics of the test-taking population. They must also have an awareness of the content students are expected to learn, and an explicit theory of how ELLs with disabilities learn in this particular content area. The theory of learning establishes the knowledge and skills that are important to assess; an understanding of the student helps educators better understand the ways in which students can show what they have learned.

The top left vertex of the triangle is observation, which includes carefully designed testing tasks that will support the kinds of inferences about student learning that educators want to make with the test data. Appropriate test administration practices, developed through

assessment training for decision makers and test administrators, accompanies a well-designed test. The top right vertex, interpretation, includes the tools and methods of reasoning about student learning. This list contains different ways of determining the quality of the assessment items (e.g., conducting Differential Item Functioning or DIF analyses on the test or on items to find out how the test functions for different groups of test takers), as well as methods of reporting test scores to the public.

To this triangle we added a center piece labeled State Assessment Policy. It is part of each of the vertices and reflects the idea that when creating a validity argument, policies must be established that reflect the way educators view student cognition, the types of assessment tasks and conditions used to measure student learning, and the way assessment data are interpreted. State assessment policies that dictate the way in which ELLs and students with disabilities are to participate in state tests (e.g., accommodations for ELLs with disabilities are decided by an IEP team that does not include English as a second language expertise) as well as how their scores are to be interpreted if they receive an accommodation (e.g., scores are reported for only the disability or ELL group) should all reflect the way the vertices of the assessment triangle interact with each other. This ideal has rarely been realized in practice, and to the extent that it has not, state validity arguments for their assessments are weakened.

Given the emphasis in school accountability efforts on students with disabilities and ELLs as separate subgroups, few states had the time to consider ELLs with disabilities when currently existing tests were developed. State personnel and test developers are not necessarily familiar with the characteristics of ELLs with disabilities and how these students best display their knowledge and skills. Therefore, assessment decision makers at the school and district level may struggle to make appropriate participation and accommodations decisions for these students.

For example, educators may exempt a student from a test if they are unaware that the student could take the test with accommodations matched to the student's language-related and disability needs. As a result, when stakeholders look at assessment scores, there may be students who are not represented in the data, thus affecting the validity of the test results. Furthermore, state assessment policies may not explicitly address assessment decisions for ELLs with disabilities, providing limited guidance for educators of these students and resulting in poor quality test participation decisions.

The IVARED project contains multiple research activities, some of which address improving educators' knowledge of the students and increasing the specific guidance found in state assessment policies. The two project activities that are the focus of this paper identify promising practices for participation, accommodation, and test score interpretation decisions for ELLs with disabilities. First, the project called on national experts to provide input on a set of principles and guidelines for assessing these students. Second, a series of online educator focus groups in the participating states obtained in-depth information on how educators make assessment participation decisions at the school and district level. Having a better understanding of the way test participation decisions are made in each state, and how processes might be affected by state-specific contexts, will support the development of a web-based training module on appropriate assessment decision making for ELLs with disabilities. The module will be used in each of the partner states to improve educators' understanding of assessment practices for ELLs with disabilities and to hopefully increase the quality of their decisions.

This paper compares and contrasts the set of expert-developed principles with the information provided by educators about actual practice. In doing so, we examine areas where

current school and district practices seem to be well-aligned with expert recommendations as well as those where improvements are needed.

Methods of Data Collection and Analysis

The data for this paper are drawn from two studies. Research methods used for each one are described separately.

Expert Principles and Guidelines

The IVARED project identified essential principles of inclusive and valid assessments for English language learners (ELLs) with disabilities. These principles were developed from a Delphi expert review process with nationally recognized experts in special education, English as a second language or bilingual education, assessment, and accountability. Additional input was obtained through presentation of the principles at national assessment and education conferences as well as during meetings of the Council of Chief State School Officers State Collaborative on Assessments and Student Standards (SCASS) groups.

The Delphi Review is a group communication technique that has been widely used to predict changes and make judgments or decisions about complex topics (Dalkey & Helmer, 1963; Howell & Kemp, 2005; Linstone & Turoff, 1975; Rowe & Wright, 1999). The purpose of the method is to reach expert “consensus” (Brill, Bishop, & Walker, 2006; Rowe & Wright, 1999), however the researchers define consensus, in an area that has little or no research base (Ziglio, 1996). Careful selection of content area experts is an important step to ensure valid results. Experts brainstorm a list of proposed changes, judgments, or decisions and then rate the set of statements. The rating process, which may occur multiple times, indicates which aspects of the topic have the most and least consensus.

Although researchers may make modifications to the basic structure, a standard Delphi Review has four important characteristics (Rowe & Wright, 1999). First, respondents remain anonymous throughout the process. Anonymity can support an open and focused exchange of ideas among experts because their opinions are not subject to group dynamics or social relationships (Clayton, 1997; Rowe & Wright, 1999). Second, reiteration of items across multiple rounds of data collection allows participants to reconsider their ratings in a nonjudgmental environment (Rowe & Wright, 1999). Third, researchers can control discussion topics and use rating systems so that the most relevant information is discussed (Rowe & Wright, 1999). Finally, group responses are statistically aggregated, usually as means (Rowe & Wright, 1999). Such analyses can provide more defensible and useable results than anecdotal data from experts' comments.

The experts recruited for this activity were individuals with in-depth knowledge of assessing and instructing ELLs with disabilities and who had the availability to participate over a two-month time period. We specifically chose participants from a variety of disciplines because they had unique knowledge bases that needed to be brought together (Liu & Anderson, 2008). For our purposes we recruited 11 experts from the fields of assessment, English as a second language or bilingual education, and special education (see Appendix A for a list). In the case where experts represent different fields, 5 to 10 participants is usually considered an appropriate number (Clayton, 1997) because it allows for unique perspectives without too complicated an analysis.

One advantage of a Delphi process is that experts can be, and often are, geographically dispersed (Clayton, 1997; Rowe & Wright, 1999). In our situation the Internet was chosen as a data collection tool specifically because our experts were scattered across the country. An

electronic Delphi allows for a faster response time and facilitation of more detailed discussion (Chou, 2002; Rotondi & Gustafson, 1996). Experts can include ideas, as well as revise them, at any time and are not limited by mailing time constraints with paper-and-pencil forms (Turoff & Hiltz, 1996). The opportunity to type responses rather than handwrite them typically leads to longer answers (Chou, 2002).

A modified three-phase Delphi process (Clayton, 1997) was used for the IVARED study.

Round one. Our 11 Delphi participants first completed a semi-structured brain storming process online where they described important considerations for improving the validity of state content assessments for ELLs with disabilities. Six areas commonly cited in U.S. Department of Education state assessment peer reviews as areas in which states either excel or need improvement, were used as themes for the brainstorming. These themes were: (a) participation decision making, (b) accommodations, (c) content standards, (d) test and item development, (e) bias and sensitivity, and (f) score reporting.

Upon completion of the round one brainstorming form, all responses were collected and grouped into sets of related statements.

Round two. During round two, expert participants saw all of the statements that had been generated and were asked to rate the importance of each one on a five-point Likert scale. They also wrote additional comments about statements. For example, they could write a question or comment about another participant's statement or a clarification of how they interpreted a particular term. Delphi studies may repeat the rating process until consensus is reached about the most important items. For our study one set of ratings was sufficient because complete consensus was not a goal due to the diversity of the participants' backgrounds. The research team decided to identify the items that were consistently rated high or low across experts.

Round three. During round three, the participants reviewed all of the individual ratings from round two as well as an average rating for each statement. Areas in which there was the most agreement were summarized. Areas in which there was the most variability in ratings were highlighted, and participants were asked to propose a specific principle for these areas of disagreement.

Principles and guidelines were created through a multi-step procedure. Statements with a high average rating (4.0-5.0) were identified and grouped together by theme. Draft principles were written based on these clusters of highly rated statements. Second, the researchers examined statements that had consistently low ratings to see whether these represented different concepts from the highly rated statements or simply additional detail about aspects of the highly-rated statements (e.g., a type of practitioner who should NOT be represented on a multi-disciplinary assessment decision-making team that was highly valued by the experts). Third, draft principles within a topic area (e.g., participation decision making) were merged into one overarching principle and several sub-principles researchers called “guidelines.” Text-based comments that related to either an overarching principle or guideline were reviewed to add detail. Finally, the draft set of principles and guidelines was presented to the state partners, to the experts, and to attendees at professional conferences for further input. Wording was adjusted based on their feedback.

Educator Focus Groups

The second research activity was a series of internet focus groups in each of the collaborating states (7 groups per state; n= 232 educators) conducted in the winter and spring of 2012. An asynchronous online format for the focus groups allowed us the flexibility to include geographically dispersed educators in rural areas and smaller school districts where a large

number of ELLs, and presumably ELLs with disabilities, may be located. Conducting the focus groups online also allowed us to engage in an important form of community building with educators who often feel isolated because of the unique population they serve.

We sent out participation invitations through state department of education networks inviting English as a second language or bilingual education teachers and program coordinators, special education teachers and program coordinators, speech-language pathologists, paraprofessionals, and district assessment coordinators to participate. Our only requirement was that they have knowledge of ELLs with disabilities and were willing to talk about assessment issues. Since the focus groups were aimed at community building and finding out needs from a variety of teacher and locations within each state, the first 50 volunteers per state were placed into groups of approximately 5-8 individuals with similar job titles.

Table 2 shows the demographic characteristics of the participants.

Table 2. Demographic characteristics of focus group participants.

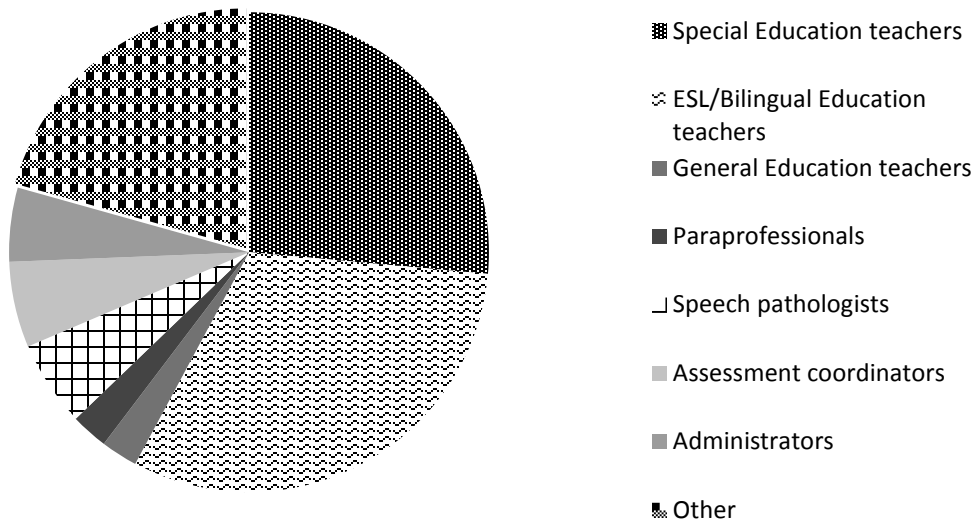
Total participants	Gender	Racial/Ethnic Background	School Locale	School Type	School Size (Number of students)
232	Male 18%	White 84%	Urban 29%	Public 90%	Under 100 7%
	Female 92%	Latino 7%	Suburban 34%	Private <1%	100-299 10%
		Asian 3%	Rural 32%	Charter 7%	300-499 28%
		Black 2%	Other 4%	Other 3%	500-699 21%
		Other 3%			700-999 16%
					1000-1999 12%
					2000-2999 2%
			3000+ 3%		

A majority of participants were white, female teachers from public schools. Even in states with a high degree of student diversity, the educators tended to fit this overall profile. School locale (e.g., urban, suburban, rural) and the school size varied. Slightly more of our participants were from suburban and rural schools with total enrollments of 300-499 students in comparison to other locations and school sizes, but there were participants from a variety of school backgrounds.

The individuals who volunteered for this study were more varied in their professional backgrounds than our initial participation criteria. We accepted volunteers from any positions if they were familiar with the topic of assessment for ELLs with disabilities. Volunteers included general education teachers, school psychologists, school principals, and a school superintendent. Widespread interest in the topic of the focus groups resulted in research study waiting lists for four of the five states. The first 50 individuals who volunteered from each state were contacted and provided with information about the study. They were invited to fill out an electronic

research consent form. If some of those 50 did not respond to our e-mails we filled in our focus groups with individuals from the waiting list. Figure 2 provides participants' job titles.

Figure 2. Focus group participants' job titles.



Special education and ESL or bilingual education teachers represented approximately equal proportions of our participants (32% and 38%, respectively). Together they made up more than half of our focus group members. The remaining percentage was split across a variety of job titles. Of our 232 total respondents, 34% (n=79) were multilingual and 66% (n=153) spoke only English.

Focus group volunteers included a sizeable number of program coordinator and administrator volunteers (n= 16; 7%). As a result, we decided to create separate focus groups for (a) individuals with classroom or small group instructional experience or service providers; and (b) individuals with test coordinator or program/school/district administration experiences. We sorted the participants from each state into groups by their job titles and then attempted to

balance the makeup of each group according to dates that participants were available to be in the study as well as their professional background.

Focus group literature (cf. Kruger & Casey, 2008) recommends developing a questioning routine of 10-14 items, beginning with broad questions to encourage participation and gradually narrowing over time. For our groups we determined that we wanted in-depth discussion and interaction between participants so we created a questioning routine of eight questions, two per day for four consecutive days. Questions were slightly different for assessment coordinators and administrators than for educators (see Appendix B for a list of questions). In keeping with the literature, our first questions for all groups related broadly to participants' experiences working with ELLs with disabilities. Over time the questions became narrower and focused specifically on making state assessment decisions for ELLs with disabilities. A former test coordinator reviewed the test coordinator/administrators questioning routine and provided feedback on how to word questions in ways these individuals could recognize.

Each focus group was conducted online via a password protected Moodle-based platform that was modified to be cleaner and simpler to use. Each group took place over a one-week time period, but the timing of groups was staggered across several months. All participants were assigned pseudonyms that were the names of trees native to their particular state. Thus, all the Arizona participants could address each other by something that sounded like a name (e.g., Saguaro, Sycamore, Juniper, etc.) which promoted interaction, while maintaining participants' anonymity. The Moodle-based program saved all of ~~participants~~participants' interactions and these transcripts (approximately 2500 pages of text) were used as the basis of our analysis.

Transcripts of focus groups were analyzed qualitatively using a semi-structured small group analysis process recommended by Krueger and Casey (personal communication,

November 1, 2012) and Krueger (1998) as an efficient, yet systematic, method of reliably coding a large amount of data. Although qualitative analysis such as thematic coding may frequently involve only one researcher and may sometimes be relatively unconcerned with issues of reliability, coding 2500 pages of text would have been a complex task beyond the limits of a single researcher (Hruschka, Schwartz, St. John, Picone-Decardo, Jenkins, & Carey, 2009). Our small group was composed of five researchers and graduate research assistants, including the focus group administrator, who all had backgrounds in education and had knowledge of ELLs with disabilities. These individuals were able to challenge each other's assumptions about the interpretation of the focus group data and arrive at a group consensus on coding major themes in the data.

Each member of the small group reviewed all of the data from the same randomly selected test coordinator/administrator group and educator group in each of the five states. The following questions were used to guide our individual review of the data as well as our semi-structured group discussion:

1. What are the groups talking a lot about?
2. What are these educators feeling about state assessments?
3. What are the most important themes related to state assessment for ELLs with disabilities?
4. What resources do these educators say they use? What do they need to improve their decision making?
5. Are there any quotes that stand out, even if they were only said once? Anything that encapsulates a lot of important ideas?

6. What assumptions are these educators making that influence their assessment decision-making?

From our review of the selected transcripts we developed a list of themes that either occurred frequently in the data or that encapsulated important ideas, even if they were mentioned infrequently. Group discussion focused on differentiating between themes mentioned in only one state or one group within a state, compared to themes mentioned across states. The lead researcher, who had also been the focus group administrator, then reviewed all of the remaining transcripts to verify the list of themes and add to the list any new themes that appeared.

Results

Principles for Assessing ELLs with Disabilities

Through the Delphi process and the subsequent rounds of feedback from professionals in the field, the research team developed five principles for assessing ELLs with disabilities on large-scale tests. We present each one with a brief explanation that also summarizes additional guidelines related to the principle. Thurlow, Liu, Ward, and Christensen (2013) contains the full discussion of each principle as well as the guidelines.

Principle One: Content standards are the same for all students.

Explanation: Standards play a key role in allocating resources and time, and in shaping students' opportunity to learn. Therefore, it is important that the same set of standards guide the instruction and assessment of all students. These standards represent the knowledge and skills students need to have to be considered proficient in specific content and to be successful after they leave school. The standards influence curricular choices and the instructional focus in classrooms. Standards also shape teaching and learning expectations and are the basis for many types of assessments. This implies that while the standards-based

performance of ELLs with disabilities may differ from the performance of the larger group of all ELLs, or all students with disabilities, the assessment outcomes should be related to a common reference point. Educators can use these data to evaluate the learning of ELLs with disabilities relative to desired goals and identify which areas of the curriculum need changes to improve student outcomes. Successful use of one common set of standards to guide the instruction and assessment of all students requires that those standards are created and written in such a way that students with a second language background and a disability can meaningfully participate. Guidelines related to Principle One discuss the importance of accessible standards written by a well-trained group of educators representing the content areas, second language instruction, and special education.

Principle Two. Test and item development include a focus on access to the content, free from bias, without changing the construct being measured.

Explanation: Valid assessment development for ELLs with disabilities should take into account students' unique characteristics. Second language learning processes are not separate from the student's disability; they interact with the disability. For example, a Chinese immigrant student who is learning English and also has a learning disability will have reading challenges that reflect a combination of his or her language processing difficulties and emerging English proficiency (e.g., limited English vocabulary, decoding text in an unfamiliar writing system). Assessments must be accessible so that every possible test taker's needs are considered and all students have equal opportunity to show their knowledge and skills. No student should be at a disadvantage when taking a test solely because he or she belongs to a particular group. At the same time, careful attention should be given to preserving the content being measured. If the content being measured is vocabulary

knowledge, providing a glossary would alter the content of the assessment. For assessment results to be valid, students' scores must reflect a measure of the intended concept without influence from irrelevant factors. The assessment should function in similar ways for all students. Guidelines related to this principle emphasize that accessible tests should incorporate the use of Universal Design principles (cf. Abedi, Kao, Leon, Mastergeorge, Sullivan, Herman & Pope, 2010; Fairbairn & Fox, 2009, Johnstone, Anderson, & Thompson, 2006; Ketterlin-Geller, 2005; Liu, & Anderson, 2008; Martiniello, 2009). These tests require the involvement of a team of experts in relevant areas of test and item development. This team needs to understand the characteristics of ELLs with disabilities who participate in the assessment. The test item development process should incorporate (a) consideration of embedded item features that may be part of newer computerized assessment formats, (b) ELLs with disabilities in item try-outs and field testing, and (c) committee-based bias reviews with experts in the content area, special education, and second language acquisition.

Principle Three. Assessment participation decisions are made on an individual student basis by an informed IEP team.

Explanation: Participation decisions refer to the in-school decisions of which test (general assessment, with or without accommodations, or alternate assessment) individual ELLs with disabilities will take. A collaborative multi-disciplinary team should make these decisions so that many different perspectives are included in the decision-making process. Participation decisions do not involve exempting students from testing. Valid assessment results for all students are necessary to ensure accountability for all student outcomes. Guidelines related to this principle discuss the importance of making individualized decisions for each student according to his or her specific needs, the importance of a diverse decision-

making team that is well-trained in assessment decision making, and the use of written policies to guide the decision-making process.

Principle Four. Accommodations for both English language proficiency (ELP) and content assessments are assigned by an IEP team knowledgeable about the individual student's needs.

Explanation: Assessment accommodations allow students to demonstrate knowledge and skills without being affected by communication issues that are unrelated to the standards being assessed. Because some students are able to show their knowledge and skills only with accommodations, providing these accommodations is essential to obtaining valid assessment results. The appropriateness of an accommodation depends on individual student needs and the construct being measured by the assessment. For example, an ELL with a learning disability may need reading supports such as having the test read to him or her, but the read aloud accommodation may be appropriate only for the math test and not for a test of reading decoding skills. Accommodations should never be assigned based solely on a student's disability category or first language. Guidelines for this principle address: (a) the use of accommodations that support the student's current level of English and native language proficiency as well as disability-related characteristics, (b) the collection of data to determine the appropriateness of a particular accommodation for a given student, (c) the development of clear policies to guide the selection of assessment accommodations for ELLs with disabilities, and (d) the need for decision-maker training on accommodations.

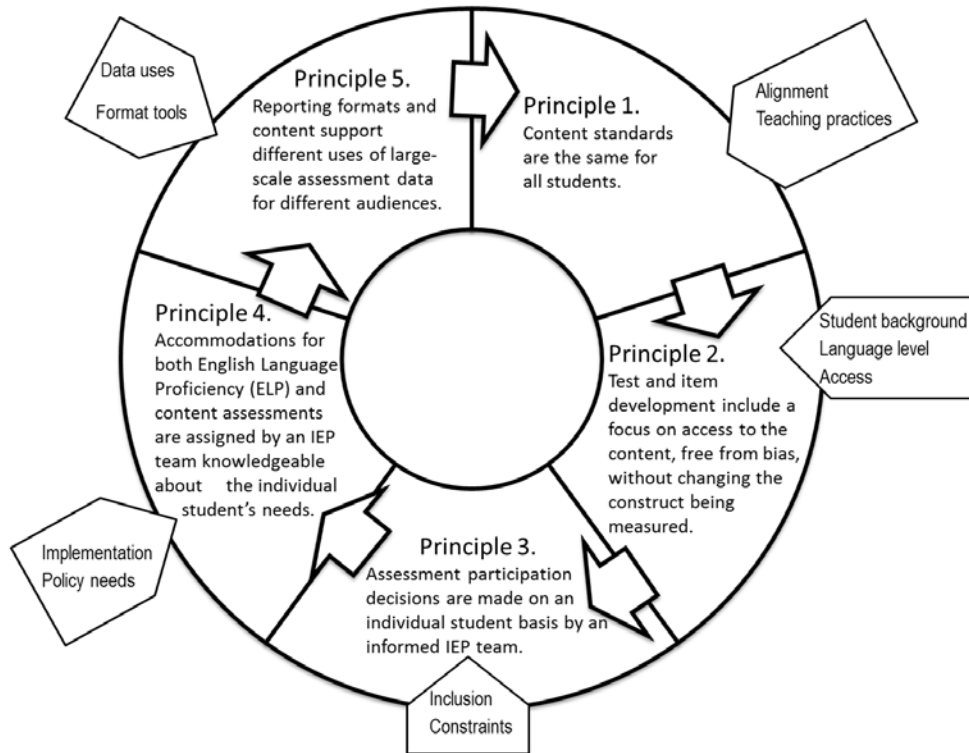
Principle Five. Reporting formats and content support different uses of large-scale assessment data for different audiences.

Student assessment data are only useful if they can be interpreted by educators and stakeholders. The appropriate use of data varies for each group that will use the data. For example, administrators may use the data for systems-level changes in their schools so they need data on all students across grade levels and years. In contrast, parents may only be interested in their own student's performance in a specific year. Providing informed and accurate descriptions of data to stakeholders in a way that contributes to their understanding will help them use the data appropriately. Additional guidelines for this principle address the need: (a) for disaggregated data that accounts for demographic differences and varying levels of students' English proficiency, (b) to highlight the characteristics of schools and districts with exceptional performance, (c) to provide guidance to educators about appropriate score interpretation and use, and (d) to use different reporting formats for different audiences.

Comparing Expert Principles to Actual School and District Practice

How do the experiences of educators at the school and district level compare to these expert-developed principles? Figure 2 highlights the areas of overlap between the five expert-developed principles for assessing ELLs with disabilities (in the circle) and comments from educator focus groups (in the arrows). In some cases, educators' perspectives addressed the exact issue described in a principle and in others, educators' comments were only somewhat related.

Figure 2. Relationship of the Themes from the Expert Principles and Guidelines Compared to Themes from the Educator Focus Groups.



Principle One. Principle One highlighted the need for content standards to be the same for all students. A few teachers were able to describe efforts that either they, as an individual teacher, or their school, had undertaken to better align the instruction ELLs with disabilities received to the grade-level standards. One participant was able to describe specific interventions, along with regular classroom assessment, that were used to target individual students' grade-level instructional needs:

“Since we are going through the entire literacy process in reading, writing and mathematics scores are closely scrutinized and teaching is more closely aligned to the standards and to students' needs. In writing, lessons are directed by needs, there are interventions in reading and mathematics and programs like Aleks (Special ed),

Pearson's tutorials or Khan Academy are used in mathematics to monitor student progress. With interventions students are tested more often (every 6 weeks) and move from group to group or continue working with the subject matter; all in small groups."

Similar to a guideline of Principle One about using a team approach for standards development, participants discussed the benefits of collaboration across disciplines given the inherent complexity of the students in considering both their English level and disability:

"The resource [special education] teacher can be helpful in many ways, but mostly with understanding behaviors. For example, there is an ELL student who is autistic that the resource teacher helps me understand the rigidity and need of routine as characteristics of autism. I rely on the assistance of the resource room teacher to also help with more social behaviors and we work together to provide English, content instruction. The resource teacher can also help with sensory issues like providing wiggle seats and arranging sensory breaks to help a few of the ELLs with disabilities. This type of input from the resource teacher is invaluable concerning assessments as well."

In addition to addressing the content standards, some teachers discussed test preparation practices they used to help students demonstrate their grade-level knowledge and skills. This area was not mentioned in the expert-developed principles as a part of ensuring test validity, But it was important to the focus group participants. The following two excerpts describe ways in which participants prepared students to be successful on state assessments.

"I have some students who do really well when testing. We work on testing taking, and core concepts in the resource classroom from the standards with me . I have three ELL students right now in resource and both are doing really well."

“I really concentrate on pre-teaching vocabulary that they will probably see in the test. I work on building as much background knowledge as possible. We also work on what kind of visuals will help them to understand the directions and concepts in the test. Since we have to cover things that might help them everyday in the classroom during testing, teaching them to draw a quick picture or remember to look in the room where the item was can also help during testing to alleviate some of the frustration and stress.”

Several educators indicated that they were not instructing ELLs with disabilities on grade-level standards and were teaching students material more than one grade level below the grade of their enrollment. This lack of alignment between the grade-level content and the student’s instructional material created difficulties in the state assessment process because the student had not learned the content being assessed. Two participants described a lack of alignment between instruction and standards:

“Though teachers use the core curriculum standards for teaching, which are on the state content assessments. However, the standards they use for ELLs with disabilities are below these students' grade level; hence their instruction does not align with the grade level assessments.”

“The problem when a student is required to take the test at grade level, but is not functioning at grade level. The results you get is that they are far below standards. We knew that before they took the test.”

Another guideline for Principle One stated that professional development on the topic of ELLs with disabilities should be ongoing. Educators concurred with this guideline, but often indicated training they wished other school staff could have. For example, one participant indicated that general educators should have more training in differentiating instruction for ELLs

with disabilities in mainstream classrooms. Participants also mentioned barriers to receiving appropriate training such as inadequate funding and a lack of offerings that were specific to the needs of ELLs with disabilities.

Principle Two. Principle Two addressed the importance of creating accessible assessments for ELLs with disabilities. The educators generally perceived a great deal of assessment bias toward this population of students. They believed that the biased tests were inaccessible to students for a variety of reasons. Typically they saw this bias coming from the process of implementing standards-based content assessments in English with students who were not proficient in the language and who lacked a particular set of mainstream cultural experiences on which test items are often based. Comments from three focus group participants highlighted this viewpoint:

“...In any assessment, classroom or state, it must be clear that language skill is not being assessed when subject matter should be. The ELL kids are given assessments in English without accommodation for language. So rather than reading skills, vocabulary or comprehension being measured, English is measured. Same goes for ELL/w disability. Twice as hard...what is being measured.”

“Challenges students have on the standardized tests are vocabulary, prior knowledge, and lack of familiarity with this type of test. Most of my students come from countries that do not test in this format ever. Additionally, the questions are many times prefaced by long narrative - even if it is just a simple calculation question. Finally, many of the questions presume cultural knowledge on the part of the student, which ELLs obviously would not have.”

“Assessments often contain culturally specific questions that we assume all students will know, but those who are learning English may not be familiar with them.”

One guideline of this Principle Two recommends considering how embedded features of items affect assessment validity such as choice options on a test with a computer format. Some focus group participants did have specific test formatting and item-level accessibility concerns that they were able to describe.

“The [state] math on the computer has also been an unexpected challenge for some of our ELL students with disabilities. The computerized voice which reads the questions on the math assessment is very difficult to understand.”

“This year many of the students including those with ELL -Disabilities did poorly on the [state] Science test. I believe one of the factors was that the test had so much reading.... Maybe the length of the Science portion could be a topic of discussion at the state level and the decision leads to some appropriate revisions.”

In some cases, suggested formatting changes had implications for assessment scoring because testing constructs would be altered if the changes were implemented. The participant below indicated that for ELLs who are blind or have low vision, oral comprehension of a text should be allowed in place of reading comprehension because students are not able to process printed text in any format.

“I would like to see that the vision teacher has the discretion on how to modify the test for each student. For example, if the student cannot read braille or large print and is an auditory learner, and all their general education is on tape/CDs then why can't the test also should be on tape, including the reading portion... a number of Ell-VI have had very

little if any [instruction] in their native countries, they do not have a reading medium so hearing the language may give us a basic knowledge of what the student knows.”

Another factor teachers mentioned when discussing test accessibility is that they tended to scaffold classroom assessments in ways that are not typically a part of large-scale assessment formats. For example, one teacher provided students with the ability to obtain partial credit on classroom tests for showing their thinking, even if that thinking was incorrect. Doing so gave her valuable instructional information about a student's strengths and weaknesses. She believed ELLs with disabilities were not able to show their thinking on standardized math assessments given by computer and thus the computerized test was not accessible.

Principle Three. Principle Three pertains to the IEP team's role in mediating individualized student assessment participation decisions. Focus group participants often did describe individualized student decisions and the IEP team members were the primary decision-makers, although English as a second language or bilingual teachers, who were sometimes not part of an IEP team, might be consulted separately. In the following excerpt, the participant did not mention the involvement of a second language learning specialist:

“The process for determining assessment participation is the same for all students with disabilities – the IEP team meets to determine the child's functioning levels and needs for accommodations.”

A guideline related to Principle Three refers to the importance of including input from a student's caregivers as part of the IEP team recommendation, even though this could have its challenges. One participant remarked:

“It is important to bring parents on board and share how they might support their student at home so teachers share ideas with parents as well... Our students are very

motivated and have lots of parental support at home...The only challenges come from the few students who do not have supportive parents: the parents who are difficult to reach via e-mail or phone, the parents who don't reinforce at home what we ask the parents to do with/for their students, etc. The most successful strategy for obtaining student involvement is relationship building with both students and parents."

Another guideline supporting this Principle Three is that decision makers should be trained to make accurate decisions that are consistent across students with similar needs. Educators expressed the need for such training :

"As an ESL teacher I would say my greatest area of need is in understanding the assessments that are done by the Special Ed. teachers who work with my students."

Even though staff acknowledged a need for training, there were often constraints to receiving it such as a lack of funding or a lack of time in the educator's schedule.

Principle Four. This principle addressed individualized accommodation decisions made by the IEP team for both state content and English proficiency assessments. Sound decision-making practices rest on clear policies that are useable by educators. However, accommodation policies were not always evident to educators or were not adequate for ELLs with disabilities. When policies are unclear, staff may have to undertake a complicated process to get guidance on appropriate decision making:

"Sometimes State guidelines aren't very clear about what accommodations ELLs with disabilities can actually have even when it's in their IEP...Any questions I still have get sent to our testing department who gets in touch with the State for clarification. It can be very time consuming particularly if you are working in a tight testing window."

One of the guidelines supporting Principle Four pertains to the need to determine appropriate accommodations for ELLs with disabilities. In doing so students need to be familiar with the accommodations, using them routinely during instruction when possible, and never using them for the first time on an assessment. In practice this guideline appeared to be inconsistently implemented. One participant described in detail his or her school's process of offering consistent instructional and assessment accommodations:

“The students are given the accommodations that are provided throughout the school year outlined in his/her IEP’s...when we are establishing accommodations for state tests, we choose those accommodations that are provided throughout the school year. It is very important to have the daily accommodations match any state or district accommodations. This provides better results because the student knows what to expect...”

In other cases participants described the difficulty of implementing consistent instructional and assessment accommodations due to issues such as time constraints:

“Many times, we are scrambling to figure out appropriate assessments and assessment accommodations for the students due the fact that our plates are already so full...”

Participants believed that implementing accommodations in the classroom, as well as on tests, requires collaboration between general education, special education, and English as a second language or bilingual teachers. One participant, commenting on his or her wish for all educators to understand the needs of ELLs with disabilities stated:

“Many [mainstream teachers] have the mindset that all accommodations and curriculum changes are the responsibility of the special education staff. Unfortunately, that is not

accurate. Teachers are responsible for those to take place, but we are always there to collaborate and consult.”

Principle Five. This principle pertained to making appropriate reporting formats useful for the different stakeholders. The way educators, including administrators and test coordinators, used the score varied. However, many participants mentioned using large-scale content assessment scores for classroom-level decisions such as intervention planning or instructional purposes. These instructional uses of accountability assessments that are typically used for programming and curriculum decisions imply that educators may not be getting the score interpretation guidance that is needed to support appropriate use.

A participant with a high school perspective described the way his or her school placed ELLs with disabilities into classes based on overall assessment performance patterns:

“When we schedule our secondary students, our scheduling teams look at overall performance at school and statewide assessments - students who are showing deficits are encouraged to schedule into co-taught core classes or classes that focus more on their skill deficits.”

Another participant described the use of state English proficiency test scores, along with content test scores, to make classroom assignments:

“My school looks at the scores to determine the level of English Proficiency along with performance on the [state content tests] to determine which class the students will be grouped into.”

Some districts use data management systems for making assessment data, along with other student data, available to educators. However, these types of reporting systems may at times

encourage teachers to use state assessment scores for classroom-level instructional planning that was not in keeping with large-scale accountability purposes :

“Our district uses a system called Viewpoint in which all students are on . On this system the students classes, grades, district and state testing results are given. Along with this system you can look at actual strands that a particular student did well on or did not do well on. Then you can compare others to see if it is a whole class problem, is a teaching gap that needs to be focused on. This system has allowed for teachers and departments to sit down and analyze the data to see where gaps of understanding are and has allowed for them to look at their curriculum so that those gaps of understanding can be looked at closer or can be re-taught. This information has been used for staff development days and for staffings to really look at students and gaps of understanding when it comes to the assessments.”

One participant acknowledged that test scores were generally used by the school for classroom planning, but were not used in this manner for ELLs with disabilities because the students were not well-assessed by state tests:

“We use the test scores from the [alternate] testing and [state content test] to support intervention groups and give teachers targeted skills to teach - However, for ELLs with IEPs, the IEP and evaluation guides the instruction for the majority of the students since we generally feel that the [State content] is not a good assessment of a student's skills and needs when they have an identified disability.”

Implications

Our broad comparison of expert-developed principles for assessing ELLs with disabilities and on-the-ground reports of actual assessment implementation and decision-making practices in

schools highlights the fact that some recommended practices are being implemented on a limited basis in certain districts and schools. School staff, particularly special education and English as a second language or bilingual teachers, have at least some understanding of how appropriate assessment practices for these students should happen. However, overall across our five states, many recommended practices are proving to be a challenge for school and district staff to carry out. There is substantial room for improvement in the validity of content assessment scores on existing tests if school and district staff receive targeted support for improving the quality and appropriateness of test participation and accommodations decisions along with test implementation practices. Here we highlight three issues that state departments of education, as well as district and school leaders, could attend to so that the scores obtained from state content tests better represent the knowledge and skills of ELLs with disabilities.

First, teachers in general appear to need substantial guidance and support in aligning students' instruction with the grade-level standards on which state content assessments are based (see Principle One). Many educators report that they are teaching content from lower grade-levels and they think assessments should measure that lower-level content. In our focus groups educators described students who entered the U.S. school system at a variety of ages. Some had limited or inconsistent formal schooling and there were large gaps in their content knowledge as a result. All of them, by definition, had limited English skills and thus had difficulty understanding grade-level content instruction. Furthermore, for those students who have a disability that affects their learning, they may need to strengthen content knowledge and skills that their peers without disabilities have learned at previous grades. When all of these circumstances are combined, teachers of ELLs with disabilities face a unique challenge to develop missing knowledge and skills from previous grades, in a language students cannot use

fluently, while still addressing grade-level expectations. Few educators have the opportunity to develop this skill set until they are on the job. They need support to improve the grade-level standards-based outcomes of this population of students.

As a useful starting point, some state departments of education have begun to explore the idea of “learning progressions” for students with significant cognitive disabilities who take alternate assessments based on alternate achievement standards (for more on this topic see the work of the National Alternate Assessment Center at <http://www.naacpartners.org>). Learning progressions “articulate the essential core concepts and processes of a discipline” and show “movement toward increased understanding” (Hess & Kearns, 2011, p. 2). For example, students with significant cognitive disabilities might be developing the skill of sequencing ideas in creating informational texts, a possible grade-level standard in English language arts, even if they are presenting information via discussion, drawings, dictation of ideas, or written text with invented spellings (Hess & Kearns, 2011). A learning progression such as this provides access points to a grade-level standard for students who may struggle to produce written text but may still be able to sequence ideas. A set of learning progressions in the tested content areas can be used for curriculum planning, classroom assessment development, and to ensure that students with learning challenges are still provided with access to the essential content standards on which state assessments are based.

The Hawaii Department of Education has taken a unique approach in creating these learning progressions for all students, including second language learners and students who do not have significant cognitive disabilities (Hess, 2012; Kurizaki, 2011). Presumably ELLs with disabilities would be included. The Hawaii learning progressions provide a model of how other state departments of education might support teachers by making the learning targets clear for a

variety of academically struggling students, but especially for ELLs with disabilities. This is a crucial first step in aligning actual classroom instruction with the grade-level standards that will be tested.

Second, educators assert that it is extremely important to have multi-disciplinary collaboration between all types of teachers and service providers who work with ELLs with disabilities in order to promote appropriate large-scale assessment decision making for a population of students with complex needs. They generally acknowledge that parents or caregivers should be included to the extent possible. This viewpoint adheres closely to Principle Three “Assessment participation decisions are made on an individual student basis by an informed IEP team.” Among the educators who participated in our focus groups, the IEP team was typically the place where individualized state assessment participation and accommodations decisions were made for ELLs with disabilities.

A variety of educators participated on IEP teams. Yet the team meetings sometimes did not include ESL or bilingual education teachers and it was unclear whether the actual general educators who worked with a specific student were members of the team. This situation occurred due to a variety of logistical constraints such as heavy teacher workloads, conflicting schedules, and varying levels of familiarity with the instructional and assessment needs of ELLs with disabilities. In addition, while parents or caregivers were generally part of the IEP team and were, in theory, present when assessment decisions were made for their student, parents may lack the background knowledge and the English skills to be equal partners in the decision-making process. IEP teams need support to create a truly multi-disciplinary team that represents all of the student’s educational experiences (e.g., general education, special education, English as a second language or bilingual education) and their home experiences. School and district staff

may need to explore ways of changing the scheduling of IEP meetings and ways of decreasing the work load of ESL and bilingual teachers in order to support IEP team participation by all of a child's teachers.

To be fully contributing members of such a team, members who are educators need training so that they all possess the skill set needed to make good decisions about state testing for a child whose learning is affected by a disability and second language acquisition. A lack of knowledge about students' needs may be one reason that IEP team membership is not representative of all of a student's educational experiences. For example, language teachers in our focus groups indicated they lacked a detailed understanding of the special education side of a student's education. Special education teachers indicated that general educators needed to understand how to provide instructional accommodations that were similar to the accommodations a child might require on a state assessment so that accommodations could be provided consistently.

Finally, teachers say that they are giving significant attention to using state content assessment scores in educational decision making, but appear to be using these test scores for classroom placement decisions and lesson planning. These are not the intended uses of large-scale accountability assessment scores. The consistency with which educators spoke about using data for classroom planning, and the level of detail they provided, suggested that school staff may not be receiving the type of guidance they need on interpreting and applying state assessment scores in an appropriate manner (see Principle Five). More detailed and specific guidance from the state education agency may support more appropriate reasoning about the knowledge and skills of ELLs with disabilities.

References

- Abedi, J., Kao, J. C., Leon, S., Mastergeorge, A. M., Sullivan, L., Herman, J., & Pope, R. (2010). Accessibility of segmented reading comprehension passages for students with disabilities. *Applied Measurement in Education, 23*(2), 168-186. doi: 10.1080/08957341003673823
- AERA, APA, NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Albus, D.A., & Thurlow, M.L. (2007). *English language learners with disabilities in state English language proficiency assessments: A review of state accommodation policies* (Synthesis Report 66). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Altman, J. R., Lazarus, S. S., Quenemoen, R. F., Kearns, J., Quenemoen, M., & Thurlow, M. L. (2010). *2009 survey of states: Accomplishments and new issues at the end of a decade of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academies Press.
- Brill, J., Bishop, M., & Walker, A. (2006). The competencies and characteristics required of an effective project manager: A Web-based Delphi study. *Educational Technology Research and Development, 54*(2), 115–140.
- Chou, C. (2002). Developing the e-Delphi system: a web-based forecasting tool for educational research. *British Journal of Educational Technology, 33*(2), 233-236.
- Clayton, M. J. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology, 17*(4), 373.

- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458-467.
- Fairbairn, S., & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers. *Educational Measurement: Issues & Practice*, 28(1), 10-24.
- Hess, K. K. (2012). *Learning progressions in K-8 classrooms: How progress maps can influence classroom practice and perceptions and help teachers make more informed instructional decisions in support of struggling learners* (Synthesis Report 87). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Howell, S., & Kemp, C. (2005). Defining early number sense: A participatory Australian study. *Educational Psychology*, 25, 555-571.
- Hruschka, D., Schwartz, D., St. John, D.C., Picone-Decaro, E., Jenkins, R., & Carey, J. (2004). Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field Methods*, (16), 307-331.
- Johnstone, C. J., Anderson, M. E., & Thompson, S. J. (2006). Universally designed assessments for ELLs with disabilities: What we've learned so far. *Journal of Special Education Leadership*, 19(1), 27-33.
- Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning, and Assessment*, 4(2).
- Koenig, J.A. (Ed.). (2002). *Reporting test results for students with disabilities and English-language learners: Summary of a workshop*. Washington, DC: National Academy Press.
- Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, DC: Council of Chief State School Officers.

- Krueger, R. (1998). *Analyzing and reporting focus group results*. Thousand Oaks, CA: SAGE Publications, Inc.
- Krueger, R. & Casey, M. (2008). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: SAGE Publications, Incorporated.
- Kurizaki, V. (2011). *Educating struggling learners: Reflections on lessons learned about curriculum, instruction, and assessment* (Synthesis Report 86). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Linstone, H., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Newark, NJ: Addison-Wesley.
- Liu, K., & Anderson, M. (2008). Universal design considerations for improving student achievement on English language proficiency tests. *Assessment for Effective Intervention*, 33(3), 167-176.
- Liu, K. Barrera, M., Thurlow, M. & Shyyan, V. (2005). *Graduation exam participation and performance (2000-2001) of English language learners with disabilities* (ELLs with Disabilities Report 2). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Liu, K., Thurlow, M., Barrera, M., Guven, K., & Shyyan, V. (2005). *Graduation Exam Participation and Performance (2000-2001) of English Language Learners with Disabilities*. National Center on Educational Outcomes.
- Marion, S., Quenemoen, R., and Kearns, J. (2006). *Inclusive assessment system options: Degree of standardization and flexibility worksheets*. Working papers from the NHEAI/NAAC Collaborative Projects.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*, 160-179.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. Pellegrino, J., Chudowsky, N., and Glaser, R., Editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Rotondi, A., & Gustafson, D. (1996). Theoretical, methodological, and practical issues arising out of the Delphi method. In M. Adler & E. Ziglio (Eds.), *Gazing into the oracle: The Delphi method and its application to social policy and public health* (pp. 3-33). Bristol, PA: Jessica Kingsley Publishers.

Rowe, G. & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting, 15*, 353-375.

Spicuzza, R., Erickson, R., Thurlow M. L., & Ruhland, A. (1996a). *Input from the field on assessing students with disabilities in Minnesota's Basic Standards Exams* (Minnesota Report No. 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Spicuzza, R., Erickson, R., Thurlow M. L., & Ruhland, A. (1996b). *Input from the field on the participation of students with limited English proficiency and students with disabilities in meeting the high standards of Minnesota's Profile of Learning* (Minnesota Report No. 10). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

- Thurlow, M. & Liu, K. (2001). Can “all” really mean students with disabilities who have limited English proficiency? *Journal of Special Education Leadership*, 14 (2), 63-71.
- Thurlow, M., Liu, K., Ward, J., & Christensen, L. (2013). *Assessment principles and guidelines for ELLs with disabilities*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M., Ysseldyke, J., & Silverstein, B. (1995). Testing accommodations for students with disabilities. *Remedial and Special Education*, 16(5), 260-270.
- Turoff, M., & Hiltz, S.R. (1996). Computer-based Delphi processes. In M. Adler & E. Ziglio (Eds.), *Gazing into the oracle: The Delphi method and its application to social policy and public health* (pp. 56-85). Bristol, PA: Jessica Kingsley Publishers.
- Zehler, A., Fleischman, H., Hopstock, P., Stephenson, T., Pendzick, M., and Sapru, S. (2003). *Descriptive study of services to LEP students and LEP students with disabilities: Volume 1 research report*. Arlington, VA: Development Associates, Inc.
- Ziglio, E. (1996). The Delphi method and its contribution to decision-making. In M. Adler & E. Ziglio (Eds.), *Gazing into the oracle: The Delphi method and its application to social policy and public health* (pp. 3-33). Bristol, PA: Jessica Kingsley Publishers.

APPENDIX A: DELPHI EXPERT REVIEW PARTICIPANTS

Delphi Participants

Jamal Abedi – Professor of Education, University of California at Davis, partner at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Leonard Baca – Professor of Education and Director of Bueno Center for Multicultural Education, University of Colorado-Boulder

Judy Elliott – Consultant; former Chief Academic Officer of the Los Angeles Unified School District

Ellen Forte – President of EdCount LLC & Director of ELL Assessment Services for the National Clearinghouse for English Language Acquisition

Barbara Gerner de Garcia – Chair and Professor of Educational Foundations and Research, Gallaudet University, Washington, D.C.

Joan Mele-McCarthy – Head of School, The Summit School, Edgewater, MD.

Marianne Perie – Senior Associate, National Center for the Improvement of Educational Assessment, Inc., Dover, NH.

Teddi Predaris – Director of the Office of Language Acquisition and Title I, Instructional Services, Fairfax County Public Schools, VA

Charlene Rivera – Director of the Center for Equity and Excellence in Education, George Washington University

Edynn Sato – Director of Research and English Language Learner Assessment, WestEd

Annette Zehler – Researcher, Center for Applied Linguistics

APPENDIX B: FOCUS GROUP QUESTIONING ROUTINES

FOCUS GROUP QUESTIONING ROUTINE FOR CLASSROOM AND SMALL GROUP EDUCATORS

Day 1

- (1) Tell us about your experience working with English language learners with disabilities.
- (2) What do you think are the top three issues educators should know about to include English language learners with disabilities in classroom assessments and state (reading, math, science) assessments?

Day 2

- (3) Describe the process your building uses to decide which state assessments (reading, math, science) ELLs with disabilities should take.
- (4) Tell us how it is decided which accommodations should be used for ELLs with disabilities for state content (reading, math, science) and English proficiency assessments.

Day 3

- 5) Give us an example of how your building has used state assessment scores to make decisions about the kinds of programs and services ELLs with disabilities receive.
- 6) Your state department of education is participating in developing an online training module to help educators make state assessment decisions for ELLs with disabilities. We would like to include examples of ELLs with disabilities for this. Please give an example of an English language learner with a disability you have worked with in the last three years, without using their real name, to highlight for the online training. In addition to the student's background, this may also include what kind of state assessments the student took; which instructional accommodations the student used in the classroom; assessment accommodations used; the accuracy of the student's test scores reflecting their knowledge and skills; and, any other pertinent information.

Day 4

- 7) What changes would you like to see made in your school about how ELLs with disabilities participate in state assessments?
- 8) Do you have any other comments about the topic of making assessment decisions for ELLs with disabilities?

FOCUS GROUP QUESTIONING ROUTINE FOR TEST COORDINATORS AND PROGRAM ADMINISTRATORS

Day 1

- (1) Tell us about your experience in assessing English language learners with disabilities.
- (2) What do you think are the top three issues educators should know about to include English language learners with disabilities in classroom assessments and state (reading, math, science) assessments?

Day 2

- (3) How do you differentiate the assessment needs of ELLs with disabilities from other ELLs or other students with disabilities?
- (4) Describe the process used by the staff in your building to make assessment participation decisions for ELLs with disabilities taking the state content (reading, math, science) and English proficiency tests in grades 3-12.

Day 3

- 5) Tell us about how accommodations are decided for state content (reading, math, science) assessments.
- 6) Give us an example of how your building or district has used state assessment scores to make decisions about the kinds of programs and services ELLs with disabilities receive.

Day 4

- 7) Your state department of education is participating in developing an online training module to help educators make state assessment decisions for ELLs with disabilities. We would like to include examples of ELLs with disabilities for this. Please give an example of an ELL with a disability you have worked with in the last three years, without using their real name, to highlight for the online training.

In addition to the student's background, this may also include, e.g., how it was decided which state content and English proficiency tests they took and which accommodations (were they the same for all the tests?) were used; whether the test scores accurately reflected their knowledge and skills; what language they took the tests; how their needs were different from students who are not ELLs with disabilities, and other information you feel is relevant.

- 8) Is there anything else you would like to add to this discussion about how to support educators in making more accurate assessment participating decisions for ELLs with disabilities?