Educational Policy
2017, Vol. 31(5) 551–583

DOI: 10.1177/0895904815613443
journals.sagepub.com/home/epx
SAGE

Large-Scale Assessment and English Language Learners With Disabilities

Kristin K. Liu[1], Jenna M. Ward[1], Martha L. Thurlow[1], and Laurene L. Christensen[1]

Abstract

This article highlights a set of principles and guidelines, developed by a diverse group of specialists in the field, for appropriately including English language learners (ELLs) with disabilities in large-scale assessments. ELLs with disabilities make up roughly 9% of the rapidly increasing ELL population nationwide. In spite of the small overall percentage of students that they represent, this group experiences significant learning and assessment challenges. In the context of successfully educating all students to high standards, it is important for state education agencies, policymakers, and local education agencies to improve achievement outcomes for these students. One of the first steps in improving test performance is to design and implement comprehensive and accessible assessment policies, and consequently assessments, that address the specific needs of ELLs with disabilities. Doing so will give them the chance to demonstrate the knowledge and skills they have, thus allowing the test results to more accurately show areas for school improvement.

Keywords

assessment, English language learners, disability

[1]University of Minnesota, Minneapolis, MN, USA

Corresponding Author:
Kristin K. Liu, National Center on Educational Outcomes, University of Minnesota, 207 Pattee Hall, 150 Pillsbury Drive SE, Minneapolis, MN 55455, USA.
Email: kline010@umn.edu

Introduction

K-12 students with disabilities often are marginalized by being placed in segregated instructional settings with lower academic expectations, and by being excluded from state assessments required for students without disabilities (National Center for Learning Disabilities, 2007). To raise expectations for students, federal special education legislation (Individuals With Disabilities Education Act [IDEA], 2004) requires educators to provide equal access to mainstream, grade-level instruction and assessment for K-12 students with disabilities in public schools. Students with disabilities are expected to be served in the least restrictive environment, which for most students is the general education classroom. Legislation also requires, to the extent possible, that students with disabilities participate, with appropriate supports and services, in the same grade-level curriculum as their peers without disabilities (Cortiella, 2006; IDEA, 2004; No Child Left Behind Act of 2001). According to the National Center for Learning Disabilities (2007),

> . . . The vast majority of students receiving special education in our nation's schools—some 85 percent—are found eligible under a disability category that in no way precludes them from—with appropriate services and supports—functioning at or above grade level or from achieving proficiency on a state's academic content standards in reading and math. (p. 8)

In addition to participating in grade-level curricula and instruction, students with disabilities also must participate in school, district, and state assessment systems along with their peers without disabilities (Cortiella, 2006; IDEA, 2004; No Child Left Behind Act of 2001). Exclusion from state accountability assessments, those designed to improve the quality and effectiveness of education systems, has been a particular challenge for students who are dually identified as both a student with a disability and an English language learner (ELL; Altman et al., 2008; Rieke, Lazarus, Thurlow, & Dominguez, 2013; U.S. Department of Education, 2014b). Federal legislation mandating state accountability testing is based on the idea that educators can create higher levels of learning for all students if schools meet certain conditions. These conditions include (a) clarity about the content students are expected to learn, (b) clarity about the content that will be assessed, and (c) use of high quality information to adjust teaching practices and obtain extra learning resources (National Research Council, 1999).

For the past two decades, work has been done on the best ways to include students with disabilities (see Spicuzza, Erickson, Thurlow, & Ruhland, 1996a, 1996b; Thurlow, Ysseldyke, & Silverstein, 1995) and ELLs (e.g., Abedi, Hofstetter, & Lord, 2004; August & Hakuta, 1997; Kieffer, Lesaux,

Rivera, & Francis, 2009; Koenig, 2002; Kopriva, 2000; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007; Spicuzza, Erickson, Thurlow, Liu, & Ruhland, 1996), as separate student groups, in large-scale assessments. Relatively little work has been done on including ELLs with disabilities who have both types of learning challenges (Albus & Thurlow, 2007; Kuti & Xu, 2012; Thurlow, Liu, Ward, & Christensen, 2013). Only a handful of states have specific policies guiding assessment practices for this group of students. The lack of policy may, in part, stem from a lack of awareness of student characteristics and limited information on best assessment practices for students with dual identification.

The authors acknowledge that there are controversies surrounding the use of standardized assessments to leverage school reform (see Darling-Hammond, 2004; Wang, Beckett, & Brown, 2006). Nevertheless, the collaborative work described in this article was intended to help state education agencies (SEAs) meet existing federal requirements. The main purposes of this article are to inform policymakers about the background and characteristics of K-12 ELLs with disabilities that are relevant to developing targeted assessment policies and to propose a set of guiding principles to inform that policy development.

Who Are ELLs With Disabilities and Why Are They a Concern?

According to Watkins and Liu (2013), estimates of the size of the population of ELLs with disabilities vary according to the source. In the 2011-2012 school year, data collected from special education programs in each state indicated that ELLs represented anywhere from 0% to 31% of students with disabilities, ages 6 through 21 (U.S. Department of Education, 2014a). These students represent about 9% of ELLs nationwide (Zehler et al., 2003), but they can be concentrated in states with particularly large ELL populations such as California, New Mexico, Nevada, Texas, and Colorado. All of these states had a concentration of 15% or more ELLs with disabilities in the special education population in the 2011-2012 school year (U.S. Department of Education, 2014a). However, the same data set shows that states that have smaller ELL populations, such as Utah, could still have relatively large percentages of ELLs with disabilities (10%). As the numbers of ELLs in general continue to increase, in some areas quite rapidly, the number of ELLs with disabilities can be expected to increase as well.

There are no systematic data collected on the incidence of particular types of disabilities among ELLs, but we can make some inferences about

the most common types of disabilities for ELLs based on the general population. Typically, among all students with disabilities, those with a specific learning disability or a speech-language impairment represent the largest student groups (U.S. Department of Education 2014a; Klingner, Artiles, & Barletta, 2006). ELLs with disabilities show a similar pattern (Zehler et al., 2003). Both ELLs with learning disabilities and those with speech-language disabilities may experience difficulties with language that are related to their disability as well as difficulties that are related to the process of learning English as a new language. It can be challenging for educators to separate the two causes of language development issues and address them separately in instruction and assessment (DeMatthews, Edwards, & Nelson, 2014; Rinaldi & Samson, 2008), making it important to examine a child's needs more holistically.

Other students with lower incidence disabilities that affect language learning might also be expected to show some interaction between their disability and second language development processes. For example, a recent immigrant student who is deaf or hard of hearing may require the use of sign language in school. However, American sign language (ASL) may be a new language for this student because he or she comes from a country that uses a different form of sign language. This student may need the support of an ASL interpreter to function in an English classroom and on tests administered in English, but his or her comprehension and language production may be affected by limited proficiency in both languages. Furthermore, if a deaf student has not been taught through a manually coded signing system that visually represents English, his or her ability to decode text may be affected not only by second language development, but also by difficulties associated with not being able to manipulate phonemes the student cannot hear. Such a student may rely on other methods of comprehending the content of text. Likewise, a student with a significant cognitive disability whose primary caretakers speak another language may comprehend his or her native language better than English. However, if he or she has limited expressive communication skills, it may be difficult for educators to determine the language in which the student is dominant.

The variability in the relationship between a student's disability and his or her second language proficiency creates unique learning and assessment needs for individual students in this population. Therefore, if educators want to have the most valid and reliable accountability data for school improvements, we must account for the interconnectedness of disability and language development when assessing children (Liu & Barrera, 2013; Shyyan et al., 2013).

Large-Scale Assessment Performance of ELLs With Disabilities

Little information is publicly available about the academic achievement of this student group or about the best ways to include them in assessments (Kuti & Xu, 2012; Thurlow, Bremer, & Albus, 2011). In part, the lack of information occurs because there is no requirement for state departments of education to disaggregate large-scale assessment results for ELLs with disabilities (Rieke et al., 2013). State departments of education may have access to assessment results for these students but may not share them publicly unless a specific request is made.

From the small amount of public information that is available, it appears that ELLs with disabilities overall are among the lowest scorers in reading, math, and science content assessments (Albus & Thurlow, 2007; Liu, Barrera, Thurlow, Guven, & Shyyan, 2005; Liu, Thurlow, Barrera, Guven, & Shyyan, 2005; Thurlow et al., 2011). These reports show that as a group, the average achievement levels of ELLs with disabilities are typically well below their non-ELL peers with disabilities and their ELL peers without disabilities. Similarly, they may not reach the more advanced levels of academic English proficiency that are expected of their ELL-only peers. Estrada (2013) documented that sixth- through ninth-grade ELLs with learning disabilities in one California school district were concentrated in the beginning, early intermediate, and intermediate levels of the state English proficiency assessment, even in secondary school, after some had spent years in U.S. public schools. In state accountability testing, test scores are intended to be a reflection of how well the educational system is meeting the needs of their students, including students with disabilities, not how capable students are of learning the content (National Research Council, 1999).

While the publicly available assessment data do indicate a cause for concern, there are three key issues with the data that may make valid score interpretations difficult. First, the data that are publicly available, especially for the state English language proficiency (ELP) assessments, may not represent the achievement of all ELLs with disabilities. ELLs with certain types of disabilities in some states may be routinely excluded from all, or some portions, of the ELP test if there is no alternate ELP assessment or the students cannot demonstrate some of the skills that are assessed (Liu et al., 2013; Rieke et al., 2013). In fact, Christensen, Albus, Liu, Thurlow, and Kincaid (2013) found that state English proficiency policies in 29 states specifically allowed ELLs with certain types of disabilities to selectively participate in only some portions of the English proficiency assessment (e.g., deaf students did not take the listening and speaking test). Excluding students from the state English

proficiency assessment has the potential to lower expectations for students' academic English development and limit the kind of English instruction they receive (Abedi, 2007; Francis & Rivera, 2007). If students do not receive academic language instruction to help them acquire the language of school, their ability to learn increasingly complex reading, math, and science content in that language may decrease (Boals et al., 2015; Hakuta, 2011; Quinn, Lee, & Valdes, 2012).

In addition, information from educators (Liu et al., 2013) indicates that in some states ELLs with certain types of disabilities (e.g., significant cognitive disabilities) may not receive the English language instruction that they need and thus do not take ELP tests. In some cases, ELLs with certain types of disabilities may be inappropriately de-identified as ELLs. De-identification ends their formal relationship with English language development programs, and the need to take yearly ELP assessments, even though the students continue to have limited proficiency in English and are still eligible for language development services (U.S. Department of Education, 2014b). Some teachers report that students' language development needs are still being addressed in the special education classroom (Liu et al., 2013). However, planning for this type of language instruction does not necessarily include second language development experts and thus may not be fully addressing students' needs. When early exit from second language instructional programs occurs, students do not take annual ELP assessments, and educational decisions made with those data will not represent the best interests of all ELLs with disabilities. Furthermore, English language development expectations for students may be lowered, which in turn minimizes students' access to academic content taught in the second language.

Second, if students have not been instructed on the grade-level content and academic language, the assessment data that are available may not support valid interpretations about the knowledge and skills ELLs with disabilities have. There are some data to suggest that ELLs with disabilities have not been instructed in the same curricula as their peers without disabilities. More than a decade ago, a national survey of ELL educators found that the curricula used to instruct ELLs with disabilities was less closely aligned with state standards than curricula for students without disabilities (Zehler et al., 2003). More recently, some educators in focus groups have stated that the results from state assessments do not always yield valuable information for ELLs with disabilities because students' academic skills are often well below grade-level standards (Liu et al., 2013). These students may be particularly likely to read below grade level and may have difficulty accessing test materials that are linguistically complex. Furthermore, the types of tasks found on state assessments are not necessarily those taught in the classroom (Liu et al.,

2013). A lack of alignment between curriculum, instruction, and assessment makes it difficult to know whether poor assessment scores indicate struggles with concepts or skills in which students have been instructed, potentially requiring a reexamination of the design of programs and services, or whether students have never been instructed in the concepts and skills that are assessed.

Third, the content and ELP assessment data that are available may not support valid interpretations about the knowledge and skills students have if participating ELLs with disabilities did not receive appropriate test accommodations. To be most effective, accommodations should simultaneously address language-based and disability-related learning challenges rather than being chosen for each type of learning challenge separately (Rogers & Christensen, 2011). However, educators report that accommodations for ELLs with disabilities may be separately assigned by the English as a second language (ESL) or bilingual education teachers for ELP tests and by the Individualized Education Program (IEP) team for content area tests. When separate accommodations decisions occur, each set of decisions reflects only part of a student's needs (Liu et al., 2013). Furthermore, not every state offers accommodations to ELLs with disabilities on all portions of state assessments, particularly on ELP assessments (Christensen et al., 2013).

When ELLs with disabilities do use accommodations, it is not always clear which ones they use. The field lacks comprehensive and clear information on accommodations actually used on state assessments. Information on accommodations used for ELP assessments is particularly infrequent in the literature, potentially reflecting a lack of clarity in the field over how to accommodate students on a test of language without changing the constructs being measured. In a 2011-2012 survey of state special education and assessment directors (Rieke et al., 2013), participants from 28 states reported that their state offered accommodations to ELLs with disabilities on all sections of the state ELP assessment. Participants from 15 states reported that their state offered accommodations only on some sections of the test. This information from staff appears to differ from the wording of existing state policies on the assessment of ELLs with disabilities. Christensen et al. (2013) found that in 2011, 37 state policies contained wording indicating that students with IEPs or Section 504 Plans could use accommodations on any segment of the ELP assessment. Eleven state policies indicated that educators could request permission to use accommodations not listed in state assessment decision-making documents (Christensen et al., 2013).

To sum up, a number of factors can cause confusion for educators who must select the best test accommodations for a given student (see Liu et al., 2013). These factors include (a) the variability in the accommodations that

are allowed for each portion of a test, (b) the lack of research available on beneficial accommodations for ELLs with disabilities, (c) the availability of selected accommodations that are not listed in policy, and (d) the difficulty of determining whether specific accommodations compromise test integrity (Bolt & Thurlow, 2004). In addition, the ease of filling out computerized IEP forms by checking boxes may lead some IEP teams to select every possible accommodation for an ELL with a disability in the belief that more accommodations cannot hurt the student's score (Kuti & Xu, 2012; Liu et al., 2013).

As the highlighted literature has shown, educators need clear assessment policies to guide decision making for ELLs with disabilities so that test results are meaningful. Yet many state department of education staff (Albus & Thurlow, 2007; Altman et al., 2008; Rieke et al., 2013) point out that the developing and implementing appropriate assessment policies for ELLs with disabilities is a continuing challenge for state policymakers.

This article describes the outcome of a 3-year collaborative effort that aimed to inform and promote valid state assessments for ELLs with disabilities. Collaborators included five state departments of education (Arizona, Maine, Michigan, Minnesota, and Washington), the National Center on Educational Outcomes (NCEO), and a diverse group of national specialists in the fields of ESL, special education, and assessment. These groups and individuals developed a set of principles and guidelines for appropriate assessment of ELLs with disabilities. The intent in creating these principles and guidelines was not only to help improve the validity of assessment systems in the five partner states but also to inform the development of inclusive assessment policies at state and local levels across the country.

We first describe the process that was used to gather input from a specialized panel and to distill a set of assessment principles and related guidelines from that input. Next, we discuss the key ideas supporting each principle, as well as relevant literature supporting those ideas, and conclude with recommendations for state and local education agencies who may be implementing the principles and related guidelines.

Method

Delphi methodology was used to investigate possible ways to improve assessment results for ELLs with disabilities (for the full report see Thurlow et al., 2013). The Delphi is a structured method that brings together a diverse group of individuals knowledgeable in relevant fields, to address a complex problem (Clayton, 1997). The Delphi is often used when there is limited research on a topic and standard analytic techniques do not fully address the research questions (Ziglio, 1996). Instead, the question of interest involves a number

of interconnected issues with policy implications for which multiple solutions are advantageous (Gupta & Clarke, 1996; Linstone & Turoff, 1975). In these cases, the Delphi method is valuable because it structures participants' interactions to yield reliable decisions while also supplying the reasoning behind the decisions (Dalkey & Helmer, 1963; Rowe & Wright, 1999). At the time this project began, there was a limited research base addressing assessment of ELLs with disabilities. Therefore, the Delphi method was well suited for developing a set of principles that could guide the field.

Five main characteristics define the structure of a Delphi study. First, Delphi studies are conducted in multiple rounds that progress from brainstorming possible problem solutions to discussing and prioritizing the most highly valued options. This process allows experts to share opinions and ideas, reflect on others' viewpoints, and adapt responses (Rowe & Wright, 1999). In this way, dissenting viewpoints are incorporated to avoid possible "groupthink" pitfalls. Second, data are collected in an asynchronous manner, where experts are able to answer when it is most convenient (Ziglio, 1996). Third, experts remain anonymous to each other throughout the study, which minimizes social dynamics that might affect an open exchange of conflicting views (Clayton, 1997). Fourth, the researchers control the topics, rating systems, and feedback provided to the experts during the Delphi. Researchers are able to tailor the focus of discussions to address the most relevant information (Rowe & Wright, 1999). Fifth, the data are analyzed using mixed-method analysis, in which responses are both qualitatively and statistically analyzed (Clayton, 1997).

Expert Panel

When selecting individuals to participate in a Delphi, Ziglio (1996) suggested that participants should be practically engaged in the field and have the capacity to explore the issues at hand. Because the topic of assessment of ELLs with disabilities spans multiple fields, we invited experts from several related fields to share their unique knowledge base. These experts came from the fields of educational assessment, special education, ESL, and bilingual education (see Table 1). Because of their diverse research expertise, a wide array of diverse opinions were represented during the Delphi rounds. Twelve experts (10 women, two men) were invited to participate in the Delphi, and 11 of them chose to participate. All 11 experts completed each round of the Delphi.

Electronic Delphi

An electronic Delphi was conducted for the current study. There are a number of benefits to conducting a Delphi study electronically. Using the

Table 1. Number of Specialists Participating in the Study, by Field of Expertise.

Field of expertise	Number of specialists
Educational assessment	5
English as a second language/bilingual education	4
Special education	2
Total	11

Note. Some specialists worked in multiple fields. The primary field is recorded in this table.

computer allows for faster access to the materials as well as faster participant response times compared with a mail-based Delphi (Chou, 2002). When the pace of the Delphi is quicker, it is easier to maintain participant interest and a more interactive discussion can take place (Rotondi & Gustafson, 1996). The expert panel is also more often comfortable typing versus handwriting long responses, so the electronic format facilitates more in-depth discussion (Chou, 2002).

For this study, experts completed the Delphi using a secure Google-based website that housed all study materials. Before the first round of the Delphi began, experts were provided background information about Delphi methodology, demographic characteristics about ELLs with disabilities, and a glossary for terms that are commonly used when discussing large-scale assessment, ELLs, or students with disabilities. Experts were provided information on Delphi methodology because Delphi best practice suggests that experts should understand the assumptions that underlie the methodology and their role in facilitating meaningful results (Linstone & Turoff, 1975; Rotondi & Gustafson, 1996; Ziglio, 1996). Because some experts were from assessment backgrounds, while others were from ESL or special education backgrounds, they were provided with a glossary of terms from the three fields.

Administration Procedures

The research team conducted the Delphi in three rounds to allow experts the chance to provide their opinions, respond to others views, and adapt their responses based upon the discussion. Three rounds took a short enough time that participant engagement remained high. Experts were given 2 weeks to complete the first round, because it was the most time intensive, and 1.5 weeks to complete the next two rounds. The researchers had 1 week in between each round to analyze the data and provide feedback for the next round.

Round 1. In the first round, experts were provided with an open-ended questionnaire that asked them to provide standards, principles, or promising practices for improving the validity of assessment results for ELLs with disabilities in each of six categories. The categories were (a) participation decision making, (b) providing accommodations, (c) development of content standards, (d) development of test items on large-scale assessments, (e) bias and sensitivity reviews of large-scale assessments for item functioning concerns, and (f) score reporting. For each category, experts were asked to consider ELP assessments and state content assessments (both regular and alternate versions).

Round 2. At the start of Round 2, experts were provided with a 5-point Likert-type questionnaire built from Round 1 responses. The responses were grouped by the original six categories and then subcategorized by themes. All Round 1 comments were included, unless they were incomplete or lacked sufficient context to be comprehensible. When experts suggested the same ideas, those responses were combined. Original wording was preserved when possible, but at times it was changed to make rating easier for other participants. Three types of changes were made. First, opinions were made into statements. Second, comments that conveyed more than one idea were divided. Last, long comments were condensed. In addition to rating each item, participants were also able to provide additional text-based comments and, as recommended by the Delphi literature (Rotondi & Gustafson, 1996), they could examine all of the original comments from Round 1.

Round 3. In Round 3, participants viewed average ratings from Round 2 as well as all of the individual ratings for each item. They could also access all Round 2 comments, grouped by section and divided into areas of agreement and disagreement. Items with large discrepancies (i.e., a mixture of ratings between 1 and 5) were categorized as areas of disagreement. Items with similar ratings (defined as items with ratings of 4 and 5) were categorized as areas of agreement. After the summaries of agreement and disagreement, participants were asked to submit a general principle that addressed just the areas of disagreement. There were between one and two areas of disagreement per category.

Data Analysis

All items on the Round 2 questionnaire with an average importance rating of 4.5 or higher were grouped into broad themes, and an overarching principle was written for each group of statements. Then comments that related to that

principle, and related highly rated items from Round 2, were used to formulate the corresponding guidelines. Because the purpose of this study was to provide a set of guiding principles to inform policy development, areas of unresolved disagreement or controversy among the experts were not included in the principles and guidelines.

Results

From the Delphi data, the research team developed five broad principles for ensuring that ELLs with disabilities are appropriately included in large-scale, standards-based content assessments (Thurlow et al., 2013). A complete list of the principles and guidelines is provided in this section. Beneath each principle, guidelines provide additional information about how to best achieve the vision that the principle describes.

Principle 1: Content standards are the same for all students.

Guideline 1A: Include individuals with knowledge of content, second language acquisition, and special education on the team that writes standards.

Guideline 1B: Design standards so they are accessible to all students, including ELLs with disabilities.

Guideline 1C: Provide ongoing professional development on implementation of standards for ELLs with disabilities to ensure high quality instruction and assessment.

Principle 2: Test and item development include a focus on access to the content, free from bias, without changing the construct being measured.

Guideline 2A: Understand the students who participate in the assessment, including ELLs with disabilities.

Guideline 2B: Involve people with expertise in relevant areas of test and item development.

Guideline 2C: Use Universal Design principles in test and item development.

Guideline 2D: Consider the impact of embedded item features and accommodations on the validity of assessment results.

Guideline 2E: Include ELLs with disabilities in item tryouts and field testing.

Guideline 2F: Conduct committee-based bias reviews for every assessment through continuous, multiphased procedures.

Principle 3: Assessment participation decisions are made on an individual student basis by an informed IEP team.

Guideline 3A: Make participation decisions for individual students rather than for groups of students.

Guideline 3B: Make assessment participations in an informed IEP team representing all instructional experiences of the student, as well as parents and students, when appropriate.
Guideline 3C: Provide the IEP team with training on assessment decision making for ELLs with disabilities.
Guideline 3D: Use written policies that specifically address the assessment of ELLs with disabilities to guide the decision-making process.
Principle 4: Accommodations for both ELP and content assessments are assigned by an IEP team knowledgeable about the individual student's needs.
Guideline 4A: Provide accommodations for ELLs with disabilities that support their current levels of English proficiency, native language proficiency, and disability-related characteristics.
Guideline 4B: Collect and examine individual student data to determine appropriate accommodations for ELLs with disabilities taking ELP and content assessments.
Guideline 4C: Develop assessment accommodations policies for ELLs with disabilities that account for the need for language-related and disability-related accommodations.
Guideline 4D: Provide decision makers with training on assessment accommodations for ELLs with disabilities.
Principle 5: Reporting formats and content support different uses of large-scale assessment data for different audiences.
Guideline 5A: Use disaggregated data for ELLs with disabilities to account for demographic and language proficiency variables.
Guideline 5B: Highlight districts and schools with exceptional performance to identify characteristics that lead to success of ELLs with disabilities.
Guideline 5C: Provide interpretation guidance to educators about ways in which large-scale assessment data can be interpreted and used for educational planning.
Guideline 5D: Provide different score report formats as guides to parents and students.

The principles and guidelines are meant to provide direction in each aspect of large-scale assessment from the creation of content standards through test development, participation decisions, and score reporting.

Discussion

As a whole, the principles and guidelines provide an overview of large-scale assessment that is relevant to all stakeholders. However, individual

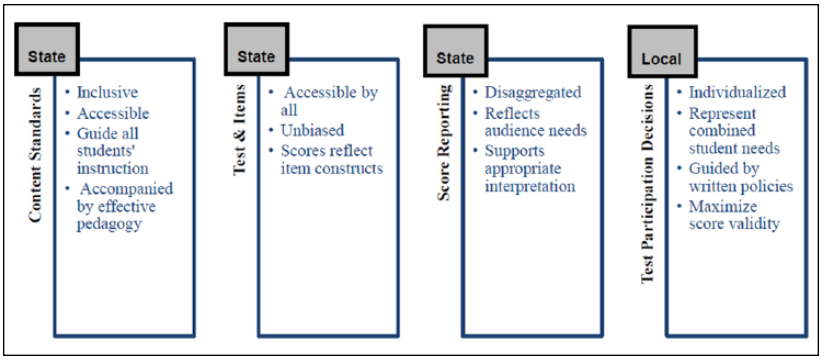

Figure 1. Key concepts contained in the *Principles and Guidelines for Assessing ELLs With Disabilities*.
Note. ELL = English language learner.

principles or guidelines may be more pertinent to particular stakeholders, as shown in Figure 1.

State departments of education may have the closest involvement with principles relating to the creation of inclusive and accessible standards, the development of assessments based on those standards, and the reporting of student results. At the local level, districts and schools are typically the sites where student assessment participation and accommodations decisions are made.

In the next section of the article, we describe how each of the principles, along with the related guidelines, relates to particular audiences, and tie the key concepts contained within them to the research literature.

State Level

At the state level, the relevant principles and guidelines (1, 2, and 5) focus on (a) providing access to, and high quality instruction in, the content that will be measured by state assessments; (b) making sure state assessments are unbiased and accessible to ELLs with disabilities; and (c) ensuring that score reporting formats and the content of reports are appropriate for the audiences who will use the information. All of these principles fall within the jurisdiction of the SEA but may be relevant to policymakers in state government as well.

Principle 1: Content standards are the same for all students. State-developed content standards represent the knowledge and skills that students at a particular

grade level should have after instruction. Although student performance on the standards, as measured by state tests, may differ, it is important that there is a common reference point which includes all students. If all students are assessed on achievement relating to the same standards, decisions about district and school programs and curricula, as well as time and resource allocation, can be made in a way that improves learning for everyone (Thurlow et al., 2013).

Current IDEA legislation emphasizes not just providing access to the general curriculum for students with disabilities but also the importance of demonstrating improved student learning outcomes with that curriculum (Hardman & Dawson, 2008). This legislation applies equally to all students with disabilities, including ELLs with disabilities. For standards to provide high quality learning opportunities, students must have well-designed instruction that is aligned to the grade-level standards (Browder, Spooner, Wakeman, Trela, & Baker, 2006; Thurlow et al., 2013). High quality learning opportunities can happen in special education and ESL or bilingual classrooms. However, experts caution that when students are segregated from the mainstream for most of the school day, their learning experiences are likely to be different from those of their mainstreamed peers (Callahan, Wilkinson, & Muller, 2010; Gándara & Orfield, 2012; Harklau, 1994; Sharkey & Layzer, 2000).

For standards-based content instruction to be high quality, special education legislation requires that it be individualized to meet the student's disability-related needs (Hoover & Patton, 2004). In addition, ELL researchers argue that instruction should be linguistically modified, if language is not the purpose of the lessons, and appropriately scaffolded so that the instruction is comprehensible to ELLs (Echevarria, Vogt, & Short, 2004; Goldenberg, 2008). Providing extensive individualization and adaptation across classroom settings requires that ESL or bilingual teachers, special education teachers, and mainstream teachers collaborate (Martin-Beltran & Peercy, 2014; Pawan & Craig, 2011) and communicate a common set of expectations to students. It also requires that the teachers are well supported in working with ELLs who have disabilities. Support can take several forms. First, the school and district must make a long-term commitment to improving the academic achievement of all students. Then within the context of that commitment, teachers receive consistent and comprehensive training on the implementation of standards with special populations of students (Lewis-Moreno, 2012), including ELLs with disabilities. Training should emphasize the core knowledge embedded in the standards, how to determine whether an individual student has learned that core knowledge, and how to remediate for missing knowledge and skills while not losing sight of grade-level content

expectations. It can also address the ways in which student learning may be affected by the combination of second language acquisition processes and a disability, and how content standards can be integrated into special education and ESL instruction (Lewis-Moreno, 2012; Thurlow et al., 2013).

All states have a set of standards in place, but as standards are revised, or new ones are developed, they must be created with all students in mind. Doing so ensures that ELLs with disabilities can meaningfully participate in standards-based instruction and assessment. Meaningful participation includes accessible standards that focus on the most important skills and that are broadly framed to avoid excluding some students whose disability or English proficiency level precludes them from demonstrating knowledge in particular ways. As an example of a standard that is framed narrowly, and which, as a result, may be difficult for ELLs with certain types of disabilities, consider the following examples from the kindergarten Common Core State Standards in English Language Arts:

> Phonological Awareness:
>
> CCSS.ELA-Literacy.RF.K2
>
> Demonstrate understanding of spoken words, syllables, and sounds (phonemes).
>
> CCSS.ELA-LITERACY.RF.K2A
>
> Recognize or produce rhyming words.
>
> CCSS.ELA-LITERACY.RF.K2.B
>
> Count, pronounce, blend, and segment syllables in spoken words. (Common Core State Standards Initiative, n.d. Available at www.corestandards.org/ELA-Literacy/RF/K/ under "Phonological Awareness")

These sample standards and benchmarks emphasize some key components of early literacy that may be difficult for some students with disabilities to demonstrate, such as those who interact with text tactilely (e.g., Braille), auditorially (e.g., talking books, screen readers), or multimodally (e.g., computerized assessments with embedded audio or video links). Being unable to show mastery of the standard means that students' scores on a related assessment may not reflect the text comprehension skills they have had the opportunity to learn through instruction with accommodations. Instead, assessment scores will reflect some limitations associated with students' disability.

Principle 2: Test and item development. Principle 2 and the related guidelines address the importance of making sure that large-scale assessment items are accessible to ELLs with disabilities. Accessible items are those that do not preclude students from showing what they have learned because of the way in which the test or item features interact with a student's second language proficiency and disability (Ketterlin-Geller, 2005, 2008; Thurlow et al., 2013). Accessibility can be addressed either through a process of universal design (Johnstone, Miller, & Thompson, 2006; Ketterlin-Geller, 2005, 2008), where items are created from the beginning to be usable by the widest variety of students possible, or through adaptation of existing items (Cawthon, Leppo, Carr, & Kopriva, 2013). In either case, test developers and SEA staff must have a clear understanding of the key concepts that are to be assessed and how ELLs with disabilities show the knowledge and skills that they have so that they can make a determination about important features of items that can, or cannot, change (Cawthon et al., 2013; Thurlow et al., 2013).

Abedi and colleagues (2011) identified five major categories of assessment accessibility features that may pose a significant barrier to students with disabilities who take standardized tests. These features include (a) cognitive features (e.g., passage type, item type, depth of knowledge, breadth), (b) textual/visual complexity (e.g., number of columns, number of pages, words per page, number of fonts, size of fonts, number of unnecessary visuals), (c) grammatical features (including passive verbs, complex verbs, relative clauses, subordinate clauses, complex noun phrases, and nominalization), (c) lexical complexity (e.g., number of long words, number of words in item or paragraph, etc.), (d) lexical density (e.g., number of unique words per item and per page or paragraph), and (e) grammatical complexity.

While these five groups of identified features could have a relatively large effect on assessment outcomes for students with disabilities, Abedi et al. (2011) found that the accessibility challenges posed by each type of feature were not necessarily the same for all (Abedi et al., 2011; see also Cawthon et al., 2013). In general, textual/visual complexity features like font style or words per page created the greatest accessibility barriers for the total group of students with disabilities. These features could be addressed relatively easily by assessment developers (Abedi et al., 2011). Many of the most challenging features they identified related to complex language use in test items. A difficulty with complex language is common to many groups of students (Abedi et al., 2011; Johnstone et al., 2006; Johnstone, Thompson, Miller, & Thurlow, 2008), but it may be difficult to determine whether the construct measured by an item should include such language. Furthermore, Abedi et al. (2011) found that language complexity, and thus the ability to take in information through reading, interacts with student characteristics in different

ways (Abedi et al., 2011). Complex language created more of an accessibility barrier, and potentially a greater difference in outcomes, for some students with disabilities compared with others.

Language-related features of assessments may also pose accessibility challenges for many second language learners as well (Cawthon et al., 2013; Liu & Anderson, 2008). For example, an expert panel on assessing ELLs with disabilities determined that a lack of concise and readable text in English proficiency assessment items was the largest accessibility barrier for these students as an overall group (Liu & Anderson, 2008).

Other possible sources of accessibility challenges not covered by Abedi et al.'s (2011) list include assessment procedures and instructions that are too complex and not intuitive to students, and an assessment that is not amenable to accommodations (Johnstone et al., 2008).

It is important for SEA staff members who create assessment policies to keep in mind that the same adaptations are not necessarily beneficial for all students with special needs across all items on a test (Cawthon et al., 2013). Students with different needs may focus on different parts of the adapted items, thus creating a variable interaction between the student and the item that may be supportive for some individuals, and not supportive for others. Therefore, steps need to be taken during the item development process to ensure that any changes to existing items have the intended effect for the desired student group (Cawthon et al., 2013). If changes to the test, such as simplifying language, result in a lowering of an items' cognitive complexity, item developers must be intentional about the way in which they make such changes (Cawthon et al., 2013).

To ensure that tests are well designed and can be accessed by the largest number of students possible, the test development team needs to have the background and training to make decisions about which aspects of items are, in fact, extraneous, and which can be altered to enhance the ways in which ELLs with disabilities interact with specific items (Thompson, Johnstone, Anderson, & Miller, 2005; Thurlow et al., 2013). Making these types of decisions entails that test developers have experience with this specific population of students.

Once items have been developed, bias review committees should examine them to determine whether there are any unforeseen sources of bias (e.g., linguistic, cultural, economic, or gender) that prevent all groups of students from interacting with items in the same way (Johnstone et al., 2008; Thompson et al., 2005; Thurlow et al., 2013). SEA staff should also field test items with a sample of students, including ELLs with disabilities, to determine whether there are additional unanticipated accessibility issues with items, or further sources of bias that need to be addressed (Johnstone et al., 2008; Thompson et al., 2005; Thurlow et al., 2013).

Principle 5: Reporting formats and content. Principle 5 addresses the need for the SEA to create clear and accessible assessment reports for a variety of audiences who are invested in the information and to support districts and schools in interpreting that information accurately. Providing disaggregated state content and English proficiency assessment data on ELLs with disabilities is crucial to improving students' educational opportunities even though disaggregated reporting for this specific group is not mandated by federal education laws (Albus, Thurlow, & Liu, 2009). The intent of sharing data publicly is to help educators, parents, and the public understand student performance and decide how to allocate resources to promote desired levels of academic achievement (Sopko & Reder, 2007). How these data are reported influences the way students are perceived and how the school system keeps track of their progress. Appropriately reported data can increase public and parental support for schools and educational programs (North Central Regional Educational Laboratory [NCREL], 1994). However, poorly reported assessment data may have the opposite effect on stakeholder attitudes and perceptions.

Most of the published recommendations for state assessment data reporting relate to the National Assessment of Educational Progress or NAEP, but these recommendations have relevance for reporting state assessment scores for individual students (Goodman & Hambleton, 2004), particularly those in special populations.

General good reporting practices include determining a few well-defined purposes for reporting assessment information prior to test administration, along with the key audiences for reports, and tailoring the reports to these individuals and purposes (Goodman & Hambleton, 2004). Each audience may have different backgrounds and require different information (Thurlow et al., 2013). Specially designed written report formats are appropriate to ensure that the audience gets the most relevant information (Thurlow et al., 2013). For example, parents may be most interested in information on their student as well as overall school and district achievement. They may also like to know what they can do to help improve their child's score (Goodman & Hambleton, 2004). Information contained within each type of report needs to be limited in scope, clearly communicated, and easily comprehended by readers who may not have much familiarity with assessment concepts (Goodman & Hambleton, 2004; Thurlow, Bremer, & Albus, 2011; Thurlow & Ives Wiley, 2006; Zenisky, Hambleton, & Sireci, 2009).

A major consideration for developing written reports is whether the readers can easily locate and extract desired information (Thurlow et al., 2008). Numerical information should be clearly presented, without the use of assessment and statistics terms such as "standard error" or "normal curve equivalent," accompanied by text that describes the data and by

visual representations of the data (Goodman & Hambleton, 2004). Organizing features such as large headings, boxes to highlight summary information, and sufficient white space help to make reports accessible (Goodman & Hambleton, 2004). For any education or testing terms that are used, a corresponding glossary of key terms helps to clearly convey information (Goodman & Hambleton, 2004). In addition, the report should provide a description of the assessment, its purpose, and how the test data will be used by schools and districts (Goodman & Hambleton, 2004). When more detail is available, it should be provided in a supplemental interpretive guide or an additional in-depth report that is disseminated at the same time as the primary report (Goodman & Hambleton, 2004).

There are not many models or research-based recommendations about the best way to report student assessment data to audiences interested in students with disabilities, especially if the students are also ELLs (Thurlow et al., 2011; Thurlow & Ives Wiley, 2006). As recently as the 2008-2009 school year, relatively few states reported the regular content or English proficiency assessment participation and performance of ELLs with disabilities (Thurlow et al., 2011). Some states did report the numbers of ELLs taking an alternate content assessment intended for students with significant cognitive disabilities (Thurlow et al., 2011).

A series of research reports written nearly a decade ago provides an example of how disaggregated content assessment data on ELLs with disabilities might be presented. The authors disaggregated data on ELLs with disabilities by language background and the student's primary disability type, when numbers of students were large enough to avoid identification of individuals (Albus, Barrera, Thurlow, Guven, & Shyyan, 2004; Albus, Thurlow, Barrera, Guven, & Shyyan, 2004; Liu, Barrera, et al., 2005; Liu, Thurlow, et al., 2005). These reports contextualized large-scale content assessment performance by providing information about the numbers of enrolled students participating in the assessments as well as students' background characteristics. What the reports did not do, was to disaggregate data on ELLs with disabilities by students' English proficiency level, a recommendation that is sometimes made in the literature (Thurlow et al., 2013).

The method of delivering information to different audiences is also important to consider. One source (NCREL, 1994) recommends that SEAs (and sometimes districts) consider communicating test information and results in multiple formats for each audience, using strategies such as briefings, meetings, written reports, press releases, and newsletter articles to share information. Providing a variety of formats, including oral delivery in English and other languages when possible, is especially important for parents of ELLs with disabilities. Some parents may have relatively

low levels of literacy in English or in their native language, which makes reliance solely on written materials a problematic communication strategy (Arias & Morillo-Campbell, 2008). Another reason that use of multiple formats is important is that parents may lack access to the Internet, a typical method used by SEAs to share assessment reports. Furthermore, it may be discouraging to any parent or community member if locating relevant information in an online assessment report takes several mouse clicks (Thurlow et al., 2011).

Local Level

When it comes to large-scale assessment, district and school-level educators make participation decisions for ELLs with disabilities, including selecting the type of test and choosing accommodations. Thus, Principles 3 and 4 apply primarily to educational policy at the local level.

Principles 3 and 4. Principle 3 and the related guidelines discuss individualized assessment participation decision making at the local level by an informed IEP team. Principle 4 and its guidelines specifically address the selection of appropriate accommodations on both content and ELP assessments. The policy implications of these two principles are similar and thus they are discussed together. Both principles share two essential components to making large-scale assessment decisions for ELLs with disabilities. First, they both advocate for individualized decisions to be made for students. Second, they both highlight the importance of a knowledgeable and collaborative IEP team in making good assessment decisions.

Individualized decisions. Underlying both principles on participation decision making is the common theme of individualized decision making. Assessment decisions for students with disabilities need to be made as part of their IEP, which is required by law to be individualized to help students with disabilities make progress in the general curriculum (IDEA, 2004). Assessment participation and accommodations decisions should not be made solely on a student's disability category, ELL status, or language background (Martinez & Humphreys, 2006). When decisions are made based upon group membership (e.g., requiring all ELLs with disabilities to take an alternate assessment, providing a translated test to all Spanish speakers regardless of students' first language literacy skills), the decision may have unforeseen detrimental effects such as lowering scores, making assessment scores difficult to interpret, and lowering expectations for the group as a whole (Abedi, 2009; Liu et al., 2013).

Depending on the assessment system of the state in which the student lives, IEP teams may have several options to choose from for content and ELP assessment participation. These options include general content assessments with, or without, accommodations, alternate content assessments based on alternate achievement standards for students with significant cognitive disabilities, and potentially alternate ELP assessments that are currently in development. Participation decisions should be informed by data collected at the individual student level. In the context of a given state's participation policies, progress-monitoring data, accomplishment of IEP goals, psychoeducational evaluations, and English, and possibly native, language proficiency data can all be used to support the selection of an appropriate assessment. Some ELLs with disabilities may not meet the criteria for participation in alternate assessments, reinforcing the need for individualized decisions. Students with disabilities should participate in challenging assessments so that they continue to receive complete access to the general curriculum (Cortiella, 2006).

Accommodations selection is another area in which individualized decisions are crucial. Due to the interaction of a student's second language development and disability, decisions that only address one area of need for the student, may inadvertently hinder the other area. For example, all ELLs in a district may be provided with a hard copy simplified English dictionary as an accommodation to help the students understand extraneous vocabulary. However, some ELLs with disabilities may have limited ability to see the print, and if their language proficiency is not sufficient to make use of the dictionary, providing it may create a distraction for certain individuals (Albus, Thurlow, Liu, & Bielinski, 2005). A hard copy ELL dictionary, if allowed for a particular content area, should only be chosen as an accommodation for students who know how to use one (previous use of the accommodation in instruction is desirable, when possible), have high enough English proficiency to comprehend entries (Albus et al., 2005; Shyyan et al., 2013), and who can interact with text in a print format. Allowing the student to use the accommodation in instruction provides the student access to instruction while giving educators the time to collect data on the utility of the accommodation (Martinez & Humphreys, 2006; Shyyan et al., 2013). Information on how the student responds to the accommodation is needed to ensure that it is not hindering the student's ability to show his or her knowledge on the assessment (Martinez & Humphreys, 2006; Shyyan et al., 2013).

Sometimes there is confusion among IEP team members as to whether the assessment accommodations listed in a student's IEP must be provided on state assessments if there is a potential that they will invalidate the assessment (e.g., a read-aloud accommodation on a reading assessment). A key

component to appropriate, individualized assessment decision making for ELLs with disabilities is the existence of clear state policies made specifically for this population of students. According to Christensen et al. (2013), several states have policy manuals that guide participation decision making for students with disabilities and ELLs separately. In some cases, those manuals may be contradictory to one another. For example, an accommodation such as reading math test items aloud may be allowed for students with disabilities, but not allowed for ELLs. In those occasions, deciding if the accommodation is allowed for an ELL with a disability can be challenging. By creating policies specifically for ELLs with disabilities, these inconsistencies can be better addressed. Policy can guide decision making while still allowing decisions to be made based upon the individual needs of the child. No policies should be put in place that make assessment participation decisions solely based on language proficiency, disability category, background characteristics, or any other group membership.

Collaboration. In addition to individualized decision making, both Principles 3 and 4 emphasize the need for collaborative decision making by all individuals involved in the student's education. Both principles presume the existence of an IEP team for ELLs with disabilities, as required by federal special education legislation. The IEP team makes assessment participation decisions for ELLs with disabilities (U.S. Department of Education, 2014b). Because the IEP is a planning tool for how services will be implemented in multiple settings, all of the relevant stakeholders should participate (Welch, 2011). Input from a variety of individuals can help to ensure that no aspect of a student's needs gets overlooked.

The key individuals on the IEP team include teachers, support staff, parents, and the student. Certain individuals are required to attend IEP meetings and be involved in assessment decisions, including general and special education teachers, parents, and a district representative (Liu & Barrera, 2013). Best practices suggest that language educators should also be part of the IEP team so that students' language development needs are appropriately accounted for in content assessment decisions (Liu & Barrera, 2013; U.S. Department of Education, 2014b), and so that appropriate decisions for accommodating annual ELP assessments can be made with special educators, rather than being made exclusively by ELL staff. In the same way, involving parents and the student themselves, when appropriate, is vitally important. When working with ELLs who have disabilities, addressing cultural and linguistic barriers to parent IEP team participation can be a significant undertaking (Lo, 2012, 2005). However, ensuring that parents understand and take an active role in the decisions being made about their child's education is necessary (Lo, 2012,

2005). Involving interpreters and cultural liaisons may help bridge barriers to communication (Lo, 2012). Other related service personnel such as school psychologists and speech pathologists can provide valuable input in collaborative assessment decision making (Welch, 2011). School psychologists are knowledgeable about assessment practices, and speech-language pathologists have expertise in language and disability services.

A lack of time and a large teaching load or caseload for educators can restrict their ability or willingness to collaborate, especially across unrelated departments (DeMatthews et al., 2014; Liu et al., 2013). District and school administrators, therefore, need to take leadership in promoting changes that support greater collaborative decision making by IEP teams. It may be necessary to reorganize teachers' schedules and classrooms and give general educators, special educators, and ELL specialists resources, as well as the time and space, to jointly plan lessons, participate on intervention teams for students who have not yet been referred to special education, and attend or provide input to IEP meetings (Martinez & Humphreys, 2006; Welch, 2011).

Additional district or school-level training may be required to ensure that educators on the IEP team understand both content and ELP standards, the possible assessment options, and how available accommodations can benefit students in both instruction and assessment (DeStefano, Shriner, & Lloyd, 2001; Rinaldi & Samson, 2008). Training on understanding second language development processes and how accommodations can support language development is also needed (Mueller, Singer, & Carranza, 2006; Rinaldi & Samson, 2008). Without these types of trainings, educators may make inappropriate test participation and accommodations decisions that can either exclude students from an assessment or prevent them from showing their true knowledge and skills (Thurlow et al., 2013). School and district administrators may also require training in the same areas so that they can support IEP team decision making (Martinez & Humphreys, 2006).

Conclusion

As state accountability systems develop and there is greater and greater use of assessment results as a basis for policy and programming decisions, it is important that assessment systems are fully inclusive. Test results must support valid inferences about the knowledge and skills of all students, including those who are dually identified as ELLs with disabilities. It is safe to assume that as the size of the total population of ELLs increases, the population of ELLs with disabilities is keeping pace. Therefore, this group of students is important for states and local education agencies to be knowledgeable about so that they can plan appropriately for ways to help students achieve grade-level standards-based outcomes.

The five principles developed through this research project are intended to create a comprehensive and unified vision for the appropriate inclusion of ELLs with disabilities in large-scale content and English proficiency assessments. The principles are also intended to maximize the validity of instructional decisions that are made based on assessment data. They are consistent with best practices recommended by the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council for Measurement in Education [NCME], 2014), *A Principled Approach to Accountability Assessments for Students With Disabilities* (Thurlow et al., 2008), and the *Accessibility Principles for Reading Assessments* (Thurlow, Laitusis et al., 2009).

The principles are not an endpoint in themselves. Rather, they should serve as a starting point for a larger, multidisciplinary conversation, along with additional research, about the appropriate instruction and assessment of ELLs with disabilities.

Acknowledgments

The authors acknowledge the significant contributions of the experts and Improving the Validity of Assessment Results for English Language Learners With Disabilities (IVARED) staff members who have worked with them in the development of these principles: *Expert Panel Members*: Jamal Abedi, Leonard Baca, Judy Elliott, Ellen Forte, Barbara Gerner de Garcia, Joan Mele-McCarthy, Marianne Perie, Teddi Predaris, Charlene Rivera, Edynn Sato, and Annette Zehler. *IVARED Staff*: Manuel Barrera, Betsy Christian, Linda Goldstone, Jim Hatten, and Hoa Nguyen.

Authors' Note

Opinions expressed herein do not necessarily reflect those of either the U.S. Department of Education or the Minnesota Department of Education.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was prepared, in part, through funds provided to the Improving the Validity of Assessment Results for English Language Learners With Disabilities (IVARED) project, which was supported by a contract (State of Minnesota Award B54419) based on a grant from the Office of Elementary and Secondary Education, U.S. Department of Education (Award S368A100011). The grant was awarded to a consortium of five state

departments of education (AZ, ME, MI, MN, WA) along with the National Center on Educational Outcomes and was led by the Minnesota Department of Education.

References

Abedi, J. (2007). English language proficiency assessment and accountability under NCLB Title III: An overview. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 3-12). Davis: University of California, Davis, School of Education.

Abedi, J. (2009). English language learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology*, *10*(2), 1-30.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*, 1-28. doi:10.3102/00346543074001001

Abedi, J., Leon, S., Kao, J., Bayley, R., Ewers, N., Herman, J., & Mundhenk, K. (2011). *Accessible reading assessments for students with disabilities: The role of cognitive, grammatical, lexical, and textual/visual features* (CRESST Report 785). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Albus, D., Barrera, M., Thurlow, M., Guven, K., & Shyyan, V. (2004). *2000-2001 participation and performance of English language learners with disabilities on Minnesota standards-based assessments* (ELLs with Disabilities Report 4). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Albus, D., & Thurlow, M. (2007). *English language learners with disabilities in state English language proficiency assessments: A review of state accommodation policies* (Synthesis Report 66). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Albus, D., Thurlow, M., Barrera, M., Guven, K., & Shyyan, V. (2004). *1999-2000 participation and performance of English language learners with disabilities on Minnesota standards-based assessments* (ELLs with Disabilities Report 1). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Albus, D., Thurlow, M., & Liu, K. (2009). *State reports on the participation and performance of English language learners with disabilities in 2006-2007* (Technical Report 54). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Albus, D., Thurlow, M., Liu, K., & Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research*, *98*, 245-256.

Altman, J., Lazarus, S., Thurlow, M., Quenemoen, R., Cuthbert, M., & Cormier, D. (2008). *2007 survey of states: Activities, changes, and challenges for special education*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Arias, M., & Morillo-Campbell, M. (2008). *Promoting ELL parental involvement: Challenges in contested times* (ERIC No. 506652). Boulder: University of Colorado, Education and Public Interest Center.

August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language minority children: A research agenda*. Washington, DC: The National Academies Press.

Boals, T., Kenyon, D., Blair, A., Cranley, M. E., Wilmes, C., & Wright, L. (2015). Transformation in K-12 English language proficiency assessment: Changing contexts, changing constructs. *Review of Research in Education*, *39*, 122-164.

Bolt, S., & Thurlow, M. (2004). Five of the most frequently allowed testing accommodations in state policy synthesis of research. *Remedial and Special Education*, *25*, 141-152.

Browder, D., Spooner, F., Wakeman, S., Trela, K., & Baker, J. (2006). Aligning instruction with academic content standards: Finding the link. *Research and Practice for Persons With Severe Disabilities*, *31*, 309-321.

Callahan, R., Wilkinson, L., & Muller, C. (2010). Academic achievement and course taking among language minority youth in U.S. schools: Effects of ESL placement. *Educational Evaluation and Policy Analysis*, *32*, 84-117. doi:10.3102/0162373709359805

Cawthon, S., Leppo, R., Carr, T., & Kopriva, R. (2013). Toward accessible assessments: The promises and limitations of test item adaptations for students with disabilities and English language learners. *Educational Assessment*, *18*, 73-98.

Chou, C. (2002). Developing the e-Delphi system: A web-based forecasting tool for educational research. *British Journal of Educational Technology*, *33*, 233-236.

Christensen, L., Albus, D., Liu, K., Thurlow, M., & Kincaid, A. (2013). *Accommodations for students with disabilities on state English language proficiency assessments: A review of 2011 state policies*. Minneapolis: University of Minnesota, National Center on Educational Outcomes, Improving the Validity of Assessment Results for English Language Learners with Disabilities.

Clayton, M. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, *17*, 373-386.

Common Core State Standards Initiative. (n.d.). English language arts standards - reading: Foundational skills - kindergarten.. Retrieved from www.corestandards.org/ELA-Literacy/RF/K/

Cortiella, C. (2006). *NCLB and IDEA: What parents of students with disabilities need to know and do*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*, 458-467.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*, 1047-1085.

DeMatthews, D., Edwards, D., & Nelson, T. (2014). Identification problems: U.S. special education eligibility for English language learners. *International Journal of Educational Research*, *68*, 27-34. doi:10.1016/j.ijer.2014.08.002

DeStefano, L., Shriner, J., & Lloyd, C. (2001). Teacher decision making in participation of students with disabilities in large-scale assessment. *Exceptional Children*, *68*, 7-22.

Echevarria, J., Vogt, M., & Short, D. (2004). *Making content comprehensible for English language learners: The SIOP model*. Needham Heights, MA: Allyn & Bacon.

Estrada, K. (2013). *Examining English language development among English language learners with specific learning disability* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (3610109).

Francis, D., & Rivera, M. (2007). Principles underlying English language proficiency tests and academic accountability for ELLs. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 13-32). Davis: University of California, Davis, School of Education.

Gándara, P., & Orfield, G. (2012). Segregating Arizona's English learners: A return to the "Mexican room"? *Teachers College Record*, *114*(9), 1-27.

Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. *American Educator*, *32*, 7-23; 42-44.

Goodman, D., & Hambleton, R. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, *17*, 145-220. doi:10.1207/s15324818ame1702_3

Gupta, U., & Clarke, R. (1996). Theory and applications of the Delphi technique: A bibliography (1975-1994). *Technological Forecasting & Social Change*, *53*, 185-211.

Hakuta, K. (2011). Educating language minority students and affirming their equal rights: Research and practical perspectives. *Educational Researcher*, *40*, 163-174.

Hardman, M., & Dawson, S. (2008). The impact of federal public policy on curriculum and instruction for students with disabilities in the general classroom. *Preventing School Failure: Alternative Education for Children and Youth*, *52*(2), 5-11. doi:10.3200/PSFL.52.2.5-11

Harklau, L. (1994). Tracking and linguistic minority students: Consequences of ability grouping for second language learners. *Linguistics and Education*, *6*, 217-244. doi:10.1016/0898-5898(94)90012-4

Hoover, J., & Patton, J. (2004). Differentiating standards-based education for students with diverse needs. *Remedial and Special Education*, *25*, 74-78.

Individuals With Disabilities Education Act [IDEA] of 2004, P.L. 108-446, 20 U.S.C. §1400 et seq.

Johnstone, C., Miller, N., & Thompson, S. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Johnstone, C., Thompson, S., Miller, N., & Thurlow, M. (2008). Universal design and multi-method approaches to item review. *Educational Measurement: Issues and Practice*, *27*, 25-36.

Ketterlin-Geller, L. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *The Journal of Technology,*

Learning and Assessment, *4*(2). Retrieved from ejournals.bc.edu/ojs/index.php/jtla/issue/archive

Ketterlin-Geller, L. (2008). Testing students with special needs: A model for understanding the interaction between assessment and student characteristics in a universally designed environment. *Educational Measurement: Issues and Practice*, *27*(3), 3-16.

Kieffer, M., Lesaux, N., Rivera, M., & Francis, D. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, *79*, 1168-1201. doi:10.3102/0034654309332490

Klingner, J., Artiles, A., & Barletta, L. (2006). English language learners who struggle with reading: Language acquisition or LD? *Journal of Learning Disabilities*, *39*, 108-128.

Koenig, J. (Ed.). (2002). *Reporting test results for students with disabilities and English-language learners: Summary of a workshop*. Washington, DC: National Academies Press.

Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, DC: Council of Chief State School Officers.

Kopriva, R., Emick, J., Hipolito-Delgado, C., & Cameron, C. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, *26*(3), 11-20. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.2007.00097.x/full

Kuti, L. M., & Xu, Y. (2012). Accommodations for English language learners on statewide English language proficiency assessment. *Journal of Special Education Apprenticeship*, *1*, 1-18.

Lewis-Moreno, B. (2012). Shared responsibility: Achieving success with English-language learners. *Phi Delta Kappa International*, *88*, 772-775.

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*, 167-176. doi:10.1177/1534508407313242

Liu, K., & Barrera, M. (2013). Providing leadership to meet the needs of English language learners with disabilities. *Journal of Special Education Administration*, *26*, 31-42.

Liu, K., Barrera, M., Thurlow, M., Guven, K., & Shyyan, V. (2005). *Graduation exam participation and performance (1999-2000) of English language learners with disabilities* (ELLs with Disabilities Report 2). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Liu, K., Goldstone, L., Thurlow, M., Ward, J., Hatten, J., & Christensen, L. (2013). *Voices from the field: Making state assessment decisions for English language learners with disabilities*. Minneapolis: University of Minnesota, Improving the Validity of Assessment Results for English Language Learners With Disabilities.

Liu, K., Thurlow, M., Barrera, M., Guven, K., & Shyyan, V. (2005). *Graduation exam participation and performance (2000-2001) of English language learners with disabilities* (ELLs with Disabilities Report 3). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Lo, L. (2005). Barriers to successful partnerships with Chinese-speaking parents of children with disabilities in urban schools. *Multiple Voices for Ethnically Diverse Exceptional Learners*, *8*, 84-95.

Lo, L. (2012). Demystifying the IEP process for diverse parents of children with disabilities. *Teaching Exceptional Children*, *44*(3), 14-20.

Martin-Beltran, M., & Peercy, M. (2014). Collaboration to teach English language learners: Opportunities for shared teacher learning. *Teachers and Teaching: Theory and Practice*, *20*, 721-737.

Martinez, R., & Humphreys, L. (2006). The best choice. *Principal Leadership*, *6*(5), 12-15.

Mueller, T., Singer, G., & Carranza, F. (2006). A national survey of the educational planning and language instruction practices for students with moderate to severe disabilities who are English language learners. *Research and Practice for Persons With Severe Disabilities*, *31*, 242-254.

National Center for Learning Disabilities. (2007). *Rewards & roadblocks: How special education students are faring under No Child Left Behind*. New York, NY: Author. Retrieved from http://www.cehd.umn.edu/NCEO/OnlinePubs/NCLD/NCLDRewardsandRoadblocks.pdf

National Research Council. (1999). *Testing, teaching, and learning: A guide for states and school districts* (Committee on Title I Testing and Assessment, R. Elmore, & R. Rothman, Eds.). Washington, DC: National Academy Press.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et seq.

North Central Regional Educational Laboratory. (1994). *Critical issue: Reporting assessment results*. Now available at http://www.rogersschools.net/common/pages/DisplayFile.aspx?itemId=3497293

Pawan, F., & Craig, D. (2011). ESL and content area teacher responses to discussions on English language learner instruction. *TESOL Journal*, *2*, 293-311.

Quinn, H., Lee, O., & Valdes, G. (2012). *Language demands and opportunities in relation to Next Generation Science Standards for ELLs*. Stanford, CA: Stanford University, Understanding Language. Retrieved from http://ell.stanford.edu/sites/default/files/pdf/academic-papers/03-Quinn%20Lee%20Valdes%20Language%20and%20Opportunities%20in%20Science%20FINAL.pdf

Rieke, R., Lazarus, S., Thurlow, M., & Dominguez, L. (2013). *2012 survey of states: Successes and challenges during a time of change*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Rinaldi, C., & Samson, J. (2008). English language learners and response to intervention. *Teaching Exceptional Children*, *40*(5), 6-14.

Rogers, C., & Christensen, L. (2011). A new framework for accommodating English language learners with disabilities (Chapter 4). In M. Russell & M. Kavanaugh

(Eds.), *Assessing students in the margin: Challenges, strategies, and techniques* (pp. 89-104). Charlotte, NC: Information Age Publishing.

Rotondi, A., & Gustafson, D. (1996). Theoretical, methodological, and practical issued 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.

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

Sharkey, J., & Layzer, C. (2000). Whose definition of success? Identifying factors that affect English language learners' access to academic success and resources. *TESOL Quarterly*, *34*, 352-368.

Shyyan, V., Christensen, L., Touchette, B., Lightborne, L., Gholson, M., & Burton, K. (2013). *Accommodations manual: How to select, administer, and evaluate use of accommodations for instruction and assessment English language learners with disabilities* (1st ed.). Washington, DC: Assessing Special Education Students and English Language Learners State Collaboratives on Assessment and Student Standards, Council of Chief State School Officers.

Sopko, K., & Reder, N. (2007). *Public and parent reporting requirements: NCLB and IDEA regulations* (In Forum). Alexandria, VA: National Association of State Directors of Special Education.

Spicuzza, R., Erickson, R., Thurlow, M., Liu, K., & Ruhland, A. (1996). *Input from the field on assessing students with limited English proficiency in Minnesota's basic requirements exams* (Minnesota Report No. 2). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

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: 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: University of Minnesota, National Center on Educational Outcomes.

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

Thurlow, M., Bremer, C., & Albus, D. (2011). *2008-09 publicly reported assessment results for students with disabilities and ELLs with disabilities* (Technical Report 59). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M., & Ives Wiley, H. (2006). A baseline perspective on disability subgroup reporting. *The Journal of Special Education*, *39*, 246-254.

Thurlow, M., Laitusis, C., Dillon, D., Cook, L., Moen, R., Abedi, J., & O'Brien, D. (2009). *Accessibility principles for reading assessments*. Minneapolis, MN: National Accessible Reading Assessment Projects.

Thurlow, M., Liu, K., Ward, J., & Christensen, L. (2013). *Assessment principles and guidelines for ELLs with disabilities*. Minneapolis: University of Minnesota, Improving the Validity of Assessment Results for English Language Learners with Disabilities.

Thurlow, M., Quenemoen, R., Lazarus, S., Moen, R., Johnstone, C., Liu, K., . . .Altman, J. (2008). *A principled approach to accountability assessments for students with disabilities* (Synthesis Report 70). Minneapolis: 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*, 260-270.

U.S. Department of Education. (2014a). 2012 IDEA Part B child count and educational environments. Retrieved from https://inventory.data.gov/dataset/8715a3e8-bf48-4eef-9deb-fd9bb76a196e/resource/a68a23f3-3981-47db-ac75-98a167b65259/download/usersshareddsf2012ideapartbchildcounteducenvrmnts.csv

U.S. Department of Education. (2014b). *Questions and answers regarding inclusion of English learners with disabilities in English proficiency assessments and Title III Annual Measurable Achievement Objectives*. Retrieved from http://www2.ed.gov/programs/sfgp/elswdfaq7182014.doc

Wang, L., Beckett, G. H., & Brown, L. (2006). Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence. *Applied Measurement in Education*, *19*, 305-328.

Watkins, E., & Liu, K. (2013, Winter/Spring). Who are English language learners with disabilities? In K. Liu, E. Watkins, D. Pompa, P. McLeod, J. Elliott, & V. Gaylord (Eds.), *Impact: Feature issue on educating K-12 English language learners with disabilities* (*Vol. 26*, No. 1). Minneapolis: University of Minnesota, Institute on Community Integration. Retrieved from https://ici.umn.edu/products/impact/261/2.html

Welch, M. (2000). Collaboration as a tool for inclusion. In S. Wade (Ed.), *Inclusive education: A casebook and readings for prospective and practicing teachers* (pp. 71-96). New York, NY: Routledge.

Zehler, A., Fleischman, H., Hopstock, P., Stephenson, T., Pendzick, M., & Sapru, S. (2003). *Descriptive study of services to LEP children and LEP children with disabilities: Summary of findings related to LEP and SpEd-LEP children* (Policy report). Arlington, VA: Development Associates.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2009). Getting the message out: An evaluation of NAEP score reporting practices with implications for disseminating test results. Applied Measurement in Education, 22(4), 359-375.

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.

Author Biographies

Kristin K. Liu, PhD, is a Research Associate at the National Center on Educational Outcomes. Her work focuses on educational accountability and large-scale assessment for K-12 English language learners (ELLs) and ELLs with disabilities.

Jenna M. Ward, MA, is a doctoral student in the School Psychology Program at the University of Minnesota. She is currently a school psychologist intern in the state of Colorado. Her past work at the National Center on Educational Outcomes addressed large-scale assessment of English language learners with disabilities.

Martha L. Thurlow, PhD, is the Director of the National Center on Educational Outcomes and Senior Research Associate at the University of Minnesota. Her work covers assessment policies, participation, accommodations, universal design, graduation exams, and alternate assessments.

Laurene L. Christensen, PhD, is a Research Associate at the National Center on Educational Outcomes. Her work focuses on a variety of assessment issues for students with disabilities, ELLs, and ELLs with disabilities. She is most interested in accommodations, alternate assessments, and leadership for school improvement.