Updated Assessment Principles and Guidelines for English Learners with Disabilities

NCEO Report 424

NCEO
National Center on Educational Outcomes

NCEO Report 424

Updated Assessment Principles and Guidelines for English Learners with Disabilities

Kristin K. Liu, Sheryl S. Lazarus, Martha L. Thurlow, Jaime Jarmin, Jenna Ward, and Laurene Christensen

August 2020

NCEO
National Center on Educational Outcomes

IDEAs that Work
U.S. Office of Special Education Programs

Project Officer: David Egnor

In collaboration with:

AEM CORPORATION CCSSO Council of Chief State School Officers NASDSE NATIONAL ASSOCIATION OF STATE DIRECTORS OF SPECIAL EDUCATION, INCORPORATED WestEd

NCEO Core Staff

Sheryl S. Lazarus, Director	Darrell Peterson
Deb A. Albus	Virginia Ressa
Jessica Bowman	Christopher Rogers
Gail Ghere	Kathy Strunk
Linda Goldstone	Martha Thurlow
Kristi K. Liu	Terri Vandercook
Charity Funfe Tatah Mentan	Yi-Chen Wu
Michael L. Moore	

Acknowledgments

Executive Summary

This report is an update of the assessment principles and guidelines for English language learners published in 2013 (Thurlow, Liu, Ward, & Christensen). That report, which was developed by the Improving the Validity of Assessment Results for English Language Learners with Disabilities (IVARED) project, presented essential principles of inclusive and valid assessments for English learners with disabilities.

Since the publication of that report, the educational context has changed. This update of the report recognizes the contextual changes, but does not alter the principles themselves.

This report presents five core principles of valid assessments for English learners with disabilities, along with a brief rationale and specific guidelines that reflect each principle. The principles are:

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

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

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

Principle 4. Accommodations for both English language proficiency (ELP) and content assessments are assigned by an IEP team knowledgeable about the individual student's needs.

Principle 5. Reporting formats and content support different uses of large-scale assessment data for different audiences.

Appendices to this report describe the Delphi data collection process, and members of the expert panel. References and selected core resources related to each principle and guideline are also included in an appendix.

Table of Contents ————————————————————

Introduction

Attention to the inclusion of students with disabilities in large-scale assessment and accountability systems emerged in the mid-1990s (Thurlow, Ysseldyke, & Silverstein, 1995). The challenge of how to include students who had not been included before (and who were sometimes targeted for exclusion) was addressed soon after by those advocating for English learners (August & Hakuta, 1997; Koenig, 2002; Kopriva, 2000). It was later that the importance of this issue was recognized for those students who were learning English and at the same time had been identified as having a disability, referred to here as English learners with disabilities (Thurlow & Liu, 2001). With the increasing numbers of these students across the nation (U.S. Department of Education, 2018; also see U.S. Department of Education, 2020), addressing their needs, and ensuring that the approaches used to include them in large-scale assessment and accountability systems, is critical.

The emphasis on including English learners with disabilities in assessments has grown out of the work that demonstrated the importance of including students with disabilities and English learners in large scale assessment systems (cf. Spicuzza, Erickson, Thurlow, Liu, & Ruhland, 1996; Spicuzza, Erickson, Thurlow, & Ruhland, 1996a, 1996b). The identified benefits grew out of the recognition that students tended to not receive needed instruction if they were not included in the large-scale assessment system, particularly the state assessment system.

Access to appropriate instruction is essential if English learners with disabilities are to progress in the curriculum and gain proficiency in English. With new and higher standards for English language arts and mathematics in the College- and Career-Ready Standards (CCRS; USDE, 2020; NGA & CCSSO, 2010), and English proficiency standards aligned to them (CCSSO, 2014), inclusion in the curriculum and appropriate standards-based instruction must be in place for English learners with disabilities. Nearly all states in the U.S. already have embraced the CCRS and have developed new standards for English language proficiency (ELP).

This report represents the collective work of a group of states committed to the appropriate inclusion of English learners with disabilities in large-scale assessment systems. These states (Minnesota as lead, Arizona, Maine, Michigan, and Washington), through the Improving the Validity of Assessment Results for English Language Learners with Disabilities project (IVA-RED), secured funding to pursue several questions related to the assessment of English learners with disabilities. One of the questions they had was how to identify the critical elements of appropriate inclusion of English learners with disabilities in large-scale assessment and accountability systems.

A set of principles and guidelines was generated using a process to systematically gather input from experts in the areas of English learners, special education, and assessment. The proce-

dures used to generate and refine these principles and guidelines, along with a description of each principle and guideline, are included in this report. (Appendix A provides a more detailed description of the Delphi procedures used to generate the basis for the principles and guidelines included in this report. Appendix B is a list of the Delphi participants.)

The principles and guidelines are meant primarily for audiences in state departments of education, especially for the leadership in assessment, special education, English learners, and those who work with them for the various purposes to which the results of large-scale assessments are put. This report is also directed to measurement experts who may sit on technical advisory committees, and testing contractors who develop large-scale assessments for system accountability. Similarly, the principles and guidelines apply to district leaders who work on district assessment systems.

The identified principles and guidelines were generated with all large-scale assessments in mind, including the general state and district assessments, the state alternate assessments based on alternate academic achievement standards (AA-AAAS), and the state ELP assessments. They also apply to alternate ELP assessments. In some cases, one or another of these assessments is targeted by a principle or guideline.

The five principles included here are meant to serve as a comprehensive and cohesive vision of ways to ensure the appropriate inclusion of English learners with disabilities in large-scale assessment systems, to make certain that their results are valid indicators of their knowledge and skills. We believe that these principles should continue to serve as a starting point for a larger, multi-disciplinary conversation about how to best assess these students. They are not the endpoint, but the beginning of a much-needed, broader discussion about the appropriate instruction and assessment of English learners with disabilities.

The guidelines under each principle provide specific information on ways to achieve the vision represented by the principle. Many of the guidelines assume that a team process is in place. This is the case for students with disabilities, but not necessarily for English learners. It is confirmed, via the guidelines, that the Individualized Educational Program (IEP) team concept is a very important one for English learners with disabilities.

Together, the principles and guidelines are intended to be consistent with the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), *Principles and Characteristics of Inclusive Assessment Systems in a Changing Assessment Landscape* developed by the National Center on Educational Outcomes (Thurlow, Lazarus, Christensen, & Shyyan, 2016), and the *Accessibility Principles for Reading Assessments* developed by the National Accessible Reading Assessment Projects (Thurlow, Laitusis, Dillon, Cook, Moen, Abedi, & O'Brien, 2009).

Although we included citations in this introduction and in the description of the Delphi process, no citations are included within the principles and guidelines themselves. This is due, in part, to the iterative Delphi process through which the principles and guidelines were derived. It is also due to the desire to keep the principles easy to read. Nevertheless, the principles and guidelines do have support in the literature. Thus, we selected some core resources related to each principle and guideline, and have included them in Appendix C.

This report includes updates that are based on recent educational policy changes since its original publication. Access to the original report may be found here: https://nceo.umn.edu/docs/OnlinePubs/ivared/IVAREDPrinciplesReport.pdf.

Overview of Principles

The five principles identified through the Delphi process each connect to the others. This is reflected in Figure 1, which shows the five principles.

Figure 1. Five Principles in the IVARED Principles and Guidelines for English Learners with
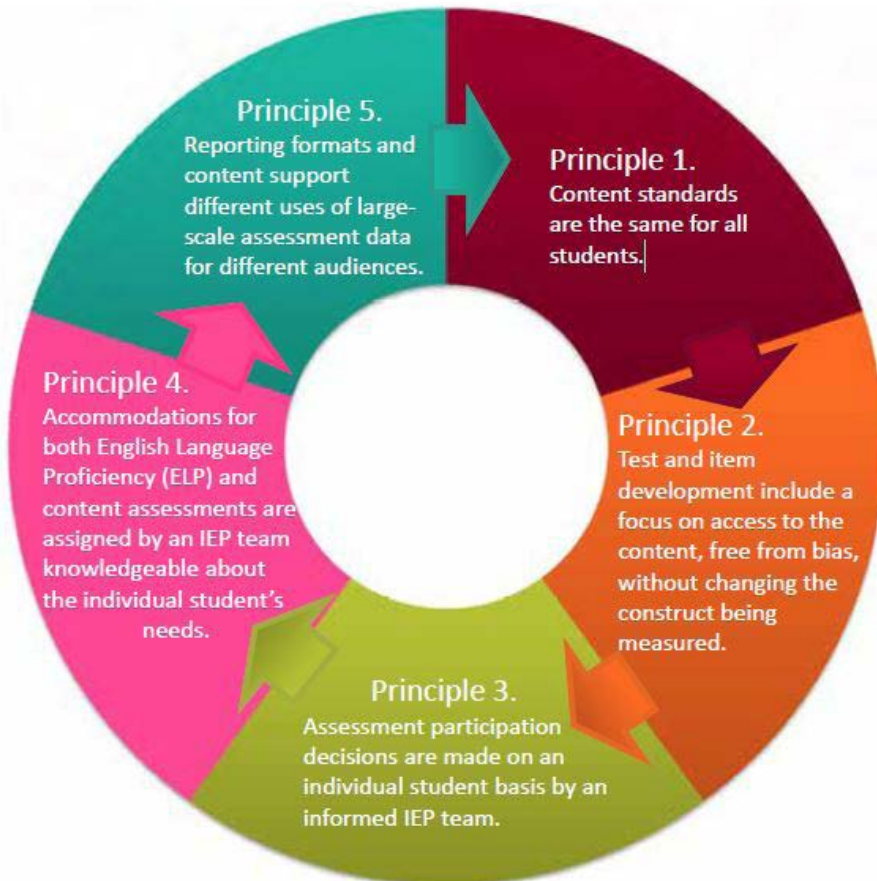
Principles and Guidelines

In this section we provide the details of each principle—what each one means in terms of specific characteristics. Rationales are provided for each principle in general, and then for each of the specific guidelines.

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

Because of the central role that standards play in allocating resources and time, and in shaping students' opportunity to learn, it is important that the same set of standards guide the instruction and assessment of all students. Content 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 educators' choice of curricula 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 English learners with disabilities may differ from the performance of the larger group of all English learners or all students with disabilities, the outcomes can be related to a common reference point. Educators can use these data to evaluate the learning of English learners with disabilities relative to desired goals and identify which areas of the curriculum need alteration to support improved student outcomes. To successfully use the same content standards with all students, the standards must be created and written in such a way that students with a second language background and a disability can meaningfully participate in the instructional and assessment processes. This principle remains important as states and consortia of states rewrite and adjust their standards over time. Three guidelines support Principle 1 (see Table 1).

Table 1. Principle 1 and Its Guidelines

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

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

A diverse standards-development team, with expertise in the content and the ways that English learners with disabilities learn that content, helps to assure that the standards are accessible to all students. Participation during standards development, rather than post-hoc, is desirable. This includes participation during initial development and during revisions and adjustments of standards over time. While some educators may lack the content-area expertise to write content standards, their perspective and experience with English learners who have disabilities are valuable.

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

From the outset, content standards should be developed to be accessible to as many students as possible. Standards should clearly focus on critical skills in which all students should be proficient on completion of their grade level, while at the same time disentangling unrelated skills that are not necessary to the performance of that standard. For example, some English learners with disabilities are not able to respond to English Language Arts (ELA) questions that require an ability to hear rhyming words because of their hearing impairments. Determining whether that skill really is important to assess is a critical first step in ensuring that standards are accessible. Depending on the decision about the importance of assessing a specific skill, it may be decided that for some students an alternative skill will need to be measured. For example, a student who is deaf might instead identify words that have comparable meanings. Similar attention has been paid to the complexity of language inferred by standards when the intent is not to test understanding of complex language.

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

Well-developed content standards are only successful at increasing standards-based learning outcomes for English learners with disabilities if they are accompanied by effective pedagogy and instructional strategies that give students access to the content. Successful teaching rests on the efficacy of teacher/leader preparation programs and continued professional development programs. Professional development should specifically address the characteristics of English learners with disabilities, ways in which these students demonstrate knowledge and skills, and how to integrate standards into the special education and English language development or bilingual education classrooms.

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

Valid assessment development for English learners with disabilities should take into account their unique characteristics. For these students, second language learning processes are not separate from the student's disability; they interact with the disability. Thus, a Chinese immigrant student who is learning English and also has a learning disability will have reading challenges that reflect a combination of 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 potential 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 while taking the assessment solely based on membership in a certain group. At the same time, careful attention should be given to preserving the content being measured. Thus, if the content being measured is vocabulary knowledge, providing a glossary would compromise the content of the assessment. For assessment results to be valid, students' scores must reflect a measure of the intended construct without influence from construct irrelevant factors. The assessment should function in similar ways for all students. Six guidelines support Principle 2 (see Table 2).

Table 2. Principle 2 and Its Guidelines

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 English learners 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 English learners with disabilities in item try-outs and field testing.

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

Guideline 2A. Understand the students who participate in the assessment, including English learners with disabilities.

To create an assessment that is valid for all students, all students should be considered while writing test items, developing test formats, and creating tests. Therefore, to create an assessment that produces valid results for English learners with disabilities, assessment developers need to have a background in both the language and disability characteristics of students. They

should also know how language learning processes may be affected by the child's disability and vice versa. In addition, test developers should consider how to write items so that they are clearly and easily interpreted by students with low incidence disabilities or who are from low frequency language groups.

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

Test and item development committees should be made up of experts in relevant areas such as psychometrics, content (e.g., reading, math, science), special education, and second language education. As appropriate, other individuals, such as parents or community members from common language groups, are included as committee members for item reviews and universal design reviews.

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

Incorporating Universal Design principles into assessment development produces test results with greater validity because more students can take the test and there may be a reduced need for accommodations. At the present time, most Universal Design research addresses content assessment accessibility issues for students with disabilities who are fluent English speakers. One important design element to consider for all students with disabilities, including English learners with disabilities, is reducing the amount of linguistic complexity where such complexity is not part of the test construct that is being measured. More research is needed about the Universal Design elements that specifically support English learners with disabilities in language learning and with their disability related needs. For example, some Universal Design considerations recommend removing distracting pictures and graphics for fluent English speakers who have disabilities, but to date, little research has addressed whether pictures and graphics help second language learners by providing additional context. Until there is a better research base specifically relating to Universal Design for English learners with disabilities, test developers will have to take the best knowledge available for students with disabilities and adapt it for language issues.

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

While developing an assessment, it is vital to consider how embedded features of items affect assessment validity for all students, including English learners with disabilities. Newer assessments that take place on the computer may allow any student to make choices about an item's appearance. These choices may either help students to accurately show what they know or hinder them from showing knowledge. For example, some computerized tests allow any student to choose the color of the font and the color of the screen background. This option may provide much needed color contrast for some English learners with low vision and allow them to read more easily. However, for some students the choice of colors may simply create a distraction.

Careful thought must be given to whether this type of an embedded feature truly provides access to the test content. In addition, English learners with disabilities must understand the embedded features so that the students can make appropriate choices.

For an online test that has embedded features, as well as for paper-pencil tests, there may still be situations in which accommodations are required to meet the needs of an individual student so that this student can meaningfully access the test. Allowable accommodations are planned from the beginning of test design because not all accessibility issues will be solvable with Universal Design principles (see Principle 4). For example, some students may need to be tested in a separate room to address their distractibility even though the test was designed from the beginning to maximize student engagement. Test developers must consider the interaction between the accommodations that English learners with disabilities will require (e.g., a screen reader for some children with learning disabilities) and the intended uses of assessment results.

Guideline 2E. Include English learners with disabilities in item try-outs and field testing.

When field testing items and new assessments, English learners with disabilities are included so that potential accessibility and bias issues that may occur with this population can be discovered. Because of the relatively small numbers of these students in some districts—and in some states—large enough samples of students may be difficult to assemble. In such a case, test developers should explore creative ways to ensure that English learners with disabilities are represented in the field-testing population. For example, states might work together to provide sufficient numbers for field testing or item tryouts.

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

For assessment results to be valid, the scores must only represent the intended construct of the assessment and no other sources of systematic error. Each assessment should be reviewed for any bias in the test that may result in unfair scoring based on group membership. A diverse group of well-trained participants with expertise in multiple areas (such as assessment, content instruction, students with disabilities, English learners, and English learners with disabilities) needs to be included in these bias reviews. Bias reviews should begin at the outset of test development and continue through each phase of creating an assessment.

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

Participation decisions refer to the in-school decisions of which test (general assessment, with or without accommodations, or alternate assessment) individual English learners with dis-

abilities will take. A team should always collaborate to make these decisions so 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. Four guidelines support Principle 3 (See Table 3).

Table 3. Principle 3 and Its Guidelines

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 participation decisions 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 English learners with disabilities.

Guideline 3D. Use written policies that specifically address the assessment of English learners with disabilities to guide the decision-making process.

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

When making participation decisions, the appropriate test should be chosen based on the student's characteristics and not the student's membership in a certain group. Deciding, for example, that all English learners with disabilities take alternate assessments would be inappropriate. Language proficiency levels and disability categories alone should not be used to justify decisions. Instead, participation decisions should be based on a team review of data collected about student characteristics to ensure valid results. (See Principle 4 for accommodations decision making.)

Guideline 3B. Make assessment participation decisions in an informed IEP team representing all instructional experiences of the student, as well as parents and students, when appropriate.

An informed IEP team includes key educators with knowledge of the student's educational background, second language acquisition status, and content learning. For English learners with disabilities, these individuals include not only general education and special education teachers, but also support staff, interpreters, psychologists, and administrators. In addition, it is important to include English as a second language, immersion, or bilingual education teachers. Any individual who can contribute unique knowledge about a student's educational experience

should be included on the team to ensure an accurate representation of the student's needs. Primary caregivers of the student are vital members of the IEP team and can provide unique insight into a student that no other member of the team can offer. The student also can offer a valuable perspective to the decision-making team, depending on age, and should be included, if feasible.

Guideline 3C. Provide the IEP team with training on assessment decision making for English learners with disabilities.

To make participation decisions that yield valid assessment results, decision makers should be trained for consistency and accuracy of decisions. They should be able to make appropriate decisions given the student's characteristics and needs, and should reach the same decisions for students with similar characteristics and needs. All members of the team need to understand the purpose of the chosen assessment and conceivable consequences of different decisions. At the district and individual school levels, important areas of expertise to be represented in training include construct relevance, psychometric issues, state guidelines, and the curriculum in which the student participates. Without this training, educators may make inappropriate test participation decisions that either exclude students from taking an assessment or that do not allow students to show their true knowledge and skills.

Guideline 3D. Use written policies that specifically address the assessment of English learners with disabilities to guide the decision-making process.

The decision-making process needs to be based on solid research with a systematic approach to choosing appropriate assessments for each student. Written policies serve to safeguard every student's right to be included in an assessment system that monitors linguistic and academic supports. State policies should address information unique to decisions made for English learners with disabilities, including types of information to be included in the decision-making process, linguistic supports that are available, and supports for students with low-incidence disabilities.

Principle 4: Accommodations for both English Language Proficiency (ELP) and content assessments are assigned by an IEP team knowledgeable about the individual student's needs.

Assessment accommodations allow students to show knowledge and skills without being affected by construct-irrelevant communication issues. Because some students are able to show their knowledge and skills only when provided 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 English learner with a learning disability may need reading supports such as hav-

ing the test read to the student by a human or through text-to-speech technology; these types of accommodations may be appropriate for the math or science test but not for a test of reading decoding skills. Accommodations should never be assigned based solely on a student's disability category or first language. Four guidelines support Principle 4 (see Table 4).

Table 4. Principle 4 and Its Guidelines

Principle 4: Accommodations for both English Language Proficiency (ELP) and content assessments are assigned by an IEP team knowledgeable about the individual student's needs.

Guideline 4A. Provide accommodations for English learners 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 English learners with disabilities taking ELP and content assessments.

Guideline 4C. Develop assessment accommodations policies for English learners 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 English learners with disabilities.

Guideline 4A. Provide accommodations for English learners with disabilities that support their current levels of English proficiency, native language proficiency, and disability-related characteristics.

When choosing accommodations for English learners with disabilities, educators should not assume students have the native language proficiency necessary to use the accommodation. Educators should consider each area of need that prevents not only the student's participation in an assessment, but also the opportunity for the student to show knowledge and skills on that assessment. English learners with disabilities may have cognitive, sensory, physical, or behavioral needs in addition to linguistic needs. The best approach for making sure all areas of need are addressed is to consider accommodations for all of these needs. Decisions should not be made on the basis of membership in a certain group (see Guideline 3A). For example, it would be inappropriate to provide a bilingual dictionary or a translated test to every student with a Hmong name without considering their proficiency in the Hmong language.

Guideline 4B. Collect and examine individual student data to determine appropriate accommodations for English learners with disabilities taking ELP and content assessments.

Data-based decisions are essential when choosing accommodations for English learners with disabilities. Before selecting an accommodation for an assessment, educators should collect data to determine the effectiveness of recommended accommodations for the student. For English learners with disabilities, understanding how their English language proficiency and disability interact is essential to choosing accommodations. In the same way, it is important to ensure that this interaction of language proficiency and disability does not affect the student's use of an accommodation. Students should be familiar with, and use regularly during instruction, the accommodations that they will use on assessments. An assessment should never be the first time a student receives an accommodation.

Guideline 4C. Develop assessment accommodations policies for English learners with disabilities that account for the need for language-related and disability-related accommodations.

Assessment accommodation policies should provide clear guidelines on both the selection and administration of individual accommodations. Policies should guide the selection of accommodations by specifying the distinction between language-related and disability-related accommodations. This could be accomplished simply by including a table of language-related needs (e.g., limited vocabulary) and disability-related needs (e.g., limited vision), and possible accommodations that address them. Policies also should define ways to ensure that the administration of accommodations results in consistent procedures across students.

Guideline 4D. Provide decision makers with training on assessment accommodations for English learners with disabilities.

Consistent test procedures that incorporate accommodations provide a way for students to show their knowledge and skills. A well-trained team of decision makers can choose and direct the administration of accommodations that will allow a student to participate in assessments in ways that produce valid results. Decision makers need to be knowledgeable about the content of the assessment, the purpose of assessment accommodations, and the relation of the accommodations to the content being assessed. The individuals administering accommodations need training in procedures that are considered to produce valid scores. The training that decision makers receive to support their decisions about participation in assessments (see Guideline 3C) may be combined with the training that they receive on assessment accommodations.

Principle 5. Reporting formats and content support different uses of large-scale assessment data for different audiences.

For student data to be useful, they need to be interpretable by educators and stakeholders. Data that are not applicable to those invested in a student's education are not worth the resources invested to collect those data. However, the appropriate use of data is different for the different audiences invested in the data. For example, administrators would like to use the data for systems level changes in their schools. In these cases, it is important that the use of the data is consistent with the purpose of the assessment when using it for educational planning. Parents, on the other hand, are most interested in their individual student. For them, understanding the results as they apply to their child's education is important. Providing informed and accurate descriptions of data to stakeholders in a way that contributes to their understanding will help all involved to use the data in an appropriate manner. Four guidelines support Principle 5 (see Table 5).

Table 5. Principle 5 and Its Guidelines

Principle 5. Reporting formats and content support different uses of large-scale assessment data for different audiences.

Guideline 5A. Use disaggregated data for English learners 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 English learners 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.

Guideline 5A. Use disaggregated data for English learners with disabilities to account for demographic and language proficiency variables.

English learners with disabilities are distinct from English learners and from students with disabilities. Data on their participation and performance should be disaggregated to allow for more meaningful interpretations of results. When numbers of students are large enough, data on English learners with disabilities should be disaggregated by level of English proficiency. When numbers are too small, disaggregated data should be reported at the next level up. For example, if reporting by proficiency level within a school is not possible, consider reporting proficient versus not proficient. If school level disaggregation is not possible, report at the district level. In addition, cross-state reporting may be helpful with states that share common assessments.

Guideline 5B. Highlight districts and schools with exceptional performance to identify characteristics that lead to success of English learners with disabilities.

Districts and schools that are performing particularly well for English learners with disabilities should be showcased at the state level. For example, the strategies of those schools where a high percentage of English learners with disabilities are making significant gains or are proficient in content areas can be promoted by a state level organization, such as the State Department of Education. In that way, other schools can learn from their success and implement changes that may help them have similar success with their students.

Guideline 5C. Provide interpretation guidance to educators about ways in which large-scale assessment data can be interpreted and used for educational planning.

School administrators and educators need to understand the ways in which large-scale assessment data can be used. For example, large-scale assessment data can be used for program evaluation, to provide a snapshot of group performance, and summative analysis. These data have limited usefulness for day-to-day classroom planning; formative sources of data are more useful for instructional purposes. Interpretation guidance will provide educators with an opportunity to use large-scale assessment data in appropriate ways.

Guideline 5D. Provide different score report formats as guides to parents and students.

When reporting large-scale assessment data of English learners with disabilities to parents and students, it is important to provide score reports that the parent and student can understand. Parents from diverse backgrounds may not have familiarity with the education system or knowledge of how large-scale assessment data are used by U.S. schools. Although not all parents or students will need a unique presentation of the data, building flexibility into the system is helpful. A variety of score formats (e.g., native language reports, face to face meetings) will help to ensure that all parents and students are informed by this educational process.

References

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

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

CCSSO (Council of Chief State School Officers). (2014). *English language proficiency (ELP) standards with correspondences to K-12 Practices and Common Core State Standards.* Retrieved from https://ccsso.org/sites/default/files/2017-11/Final%204_30%20ELPA21%20 Standards%281%29.pdf

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.

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.

NGA (National Governors' Association) and CCSSO (Council of Chief State School Officers). (2010). *Common core state standards.* Available at www.corestandards.org.

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

Thurlow, M. L., Lazarus, S. S., Christensen, L. L., & Shyyan, V. (2016). *Principles and characteristics of inclusive assessment systems in a changing assessment landscape* (NCEO Report 400). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

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

Thurlow, M. L., Quenemoen, R. F., Lazarus, S. S., Moen, R. E., Johnstone, C. J., Liu, K. K., Christensen, L. L., Albus, D. A., & Altman, J. (2008). *A principled approach to accountability assessments for students with disabilities* (Synthesis Report 70). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Ysseldyke, J. E., & Silverstein, B. (1995). Testing accommodations for students with disabilities. *Remedial and Special Education, 16*(5), 260-270.

U.S. Department of Education. (2018). *Our nation's English learners: What are their characteristics?* Retrieved from https://www2.ed.gov/datastory/el-characteristics/index.html.

U.S. Department of Education. (2020). IDEA section 618 data products: State level data files (Part B, Child count and educational environments). Retrieved from https://www2.ed.gov/programs/osepidea/618-data/state-level-data-files/index.html#part-b-menu

Appendix A

Delphi Expert Review Procedures

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) in an area that has little or no research base (Ziglio, 1996). A Delphi review is most often used when an issue is tied to a number of consequences and policy options within the field and an in-depth examination and discussion of each option is needed (Linsoten & Turoff, 1975; Turoff & Hiltz, 1996). A standard Delphi Review typically starts with the identification of a panel of experts in the topic to be discussed. Careful selection of experts is an important step to ensure valid results.

The experts recruited for this activity were individuals with in-depth knowledge of assessing and instructing English learners with disabilities who had the willingness to participate over a two-month time period. Sometimes, as was the case for the IVARED study, these experts are from diverse but related fields, and they have unique knowledge bases that need to be brought together (Liu & Anderson, 2008). For the IVARED Delphi, 11 experts were recruited from educational assessment, special education and English as a second language or bilingual education. In the case where experts represent different fields, 5 to 10 participants is an appropriate number (Clayton, 1997) as it allows for unique perspectives without too complicated an analysis. Often these experts are geographically dispersed (Clayton, 1997; Rowe & Wright, 1999).

In the IVARED study, the Internet was chosen as a data collection tool because experts lived in different parts of 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 (Turoff & Hiltz, 1996). The opportunity to type responses rather than handwrite them typically leads to longer answers (Chou, 2002).

Characteristics of a Delphi Review

A standard Delphi Review has four important characteristics (Rowe & Wright, 1999). First, respondents remain anonymous throughout the process (Clayton, 1997). 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 dis-

cussed (Rowe & Wright, 1999). Finally, group responses are statistically aggregated, usually as means (Rowe & Wright, 1999). Such analyses can provide more defensible and valid results than simply using anecdotal data from experts' comments.

The standard Delphi procedures can be modified depending on the purpose of the review (Brill et al., 2006; Linstone & Turoff, 1975; Murray & Hammons, 1995). For example, the first stage can be more structured, the number of rounds of data collection can be varied, and respondents can be asked for types of answers other than a Likert-type rating. If consensus is desired on an already established list of items, the first stage may often be omitted entirely.

A three-phase Delphi process is common (Clayton, 1997) and was used for the IVARED study. This process is described below with examples of how it was adapted for this specific research activity.

Phase 1

The first phase is relatively unstructured and involves written answers to an open-ended prompt. We chose to ask experts to comment on assessment validity topics that were taken from U.S. Department of Education assessment peer review documents. These topics included: assessment participation decision making, accommodations, content standards, test and item development, test bias and sensitivity, and score reporting. A list of key points is generated from the written answers.

Phase 2

The second phase includes at least one opportunity for experts to rate the importance or desirability of the key points generated in phase 1. A 5- or 7- point Likert-type scale is commonly used for ratings. In some studies ratings are repeated until a pre-established indicator of consensus is reached (Rotondi & Gustafson, 1996). However, for the IVARED study complete consensus was not a goal because of the diversity of the participants' backgrounds. IVARED researchers resolved to identify the items that were consistently rated high or low across experts. Thus, one set of ratings was sufficient for our purposes. Throughout the second phase of a Delphi, participants see summaries of the ratings and may also see comments made by other participants. They are given an opportunity to change their ratings based on other experts' responses.

Phase 3

In the third phase, Delphi facilitators determine what represents consensus on the rated statements and indicate which statements have the strongest degree of consensus. For the IVARED Delphi review, the research team identified all statements that were deemed important by experts. Importance was reflected by a mean rating of 4 or higher on a scale of 0 to 5, signifying

no importance to important. Items which had a mean rating of 1 or less than 1 were identified on the other end of the scale.

Delphi Expert Review Procedures References

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.

Howell, S., & Kemp, C. (2005). Defining early number sense: A participatory Australian study. *Educational Psychology, 25*, 555–571.

Linstone, H.A., & 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.

Murray, J., & Hammons, J. (1995). Delphi: A versatile methodology for conducting qualitative research. *Review of Higher Education, 18*, 423–436.

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., 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, MN: 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, 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.

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.

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 B

Delphi Participants (Positions may have changed since the Delphi was conducted)

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 C

Core Resources for Principles and Guidelines

We include here a set of core resources for each principle. The principles and guidelines themselves are based on a body of literature that includes public policy, assessment standards, and studies. The core resources that we include here are not exhaustive but are intended to provide a core body of work that reflects the intent of the principles and guidelines.

Principle 1

Alliance for Excellent Education. (2012). *The role of language and literacy in college- and career-ready standards: Rethinking policy and practice in support of English language learners* (Policy Brief). Washington, DC: Author.

Bailey, A. L., & Carroll, P. E. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education*, *39*(1), 253-294.

Gottlieb, M. (2012). Implementing the common core state standards in districts with English language learners: What are school boards to do? *The State Education Standards, 12*(2), 63-65.

Lee, O. (2019). Aligning English language proficiency standards with content standards: Shared opportunity and responsibility across English learner education and content areas. *Educational Researcher*, *48*(8), 534-542.

McLaughlin, M. J. (2012). Access for all: Six principles for principals to consider in implementing CCSS for students with disabilities. *Principal,* 22-26.

Pompa, D., & Hakuta, K. (2012). Opportunities for policy advancement for ELLs created by the new standards movement. *Understanding Language: Language, Literacy, and Learning in the Content Areas*. Stanford, CA: Stanford University.

Thurlow, M. L., & Quenemoen, R. F. (2012). Opportunities for students with disabilities from the common core standards. *The State Education Standard, 12*(2), 56-62.

Principle 2

Abedi, J. (2014). English language learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology*, *10*(2), 1-30.

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

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.

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).

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. K., Lazarus, S., Thurlow, M. L., Stewart, J., & Larson, E. (2020). *A summary of the research on test accommodations for English learners and English learners with disabilities: 2010-2018.* Minneapolis, MN: University of Minnesota, Improving Instruction for English Learners through Improved Accessibility Decisions.

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

National Center on Educational Outcomes [NCEO]. (2011, March). *Don't forget accommodations! Five questions to ask when moving to technology-based assessments* (NCEO Brief #1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Rios, J. A., Ihlenfeldt, S. D., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence-based? A multistudy systematic review and meta-analysis. Advance online publication. *Educational Measurement: Issues and Practice.* Doi:10.1111/emip.12337.

Rogers, C., & Christensen, L. (2011). A new framework for accommodating English language learners with disabilities. In M. Russell & M. Kavanaugh (Eds.), *Assessing students in the margin: Challenges, strategies, and techniques* (pp. 89-104). Charlotte, NC: Information Age Publishing.

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education*, *39*(1), 215-252.

Thurlow, M. L., Liu, K. K., Lazarus, S. S., & Moen, R. E. (2005). *Questions to ask to determine how to move closer to universally designed assessments from the very beginning, by addressing the standards first and moving on from there.* Minneapolis, MN: University of Minnesota, Partnership for Accessible Assessments (PARA). Available at http://www.readingassessment. info/resources/publications/QuestionstoAskUniversallyDesignedAssessments.pdf.

Zieky, M. J. (2015). Developing fair tests. In *Handbook of test development* (pp. 97-115). London: Routledge.

Principle 3

Elliott, J. L., & Thurlow, M. L. (2006). Addressing the needs of IEP/ELLs. *Improving test performance of students with disabilities … on district and state assessments* (2nd ed.). Thousand Oaks, CA: Corwin Press.

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.

Improving Instruction. (2020). *Improving instruction for English learners through accessibility decision making (Improving Instruction): Parent-educator toolkit.* Retrieved from https://nceo. info/About/projects/improving-instruction/parent-educator-toolkit.

Improving Instruction. (2020). *Improving instruction for English learners through accessibility decision making (Improving Instruction): Training module.* Retrieved from https://nceo.info/ About/projects/improving-instruction/training-module

Klingner, J., & Harry, B. (2006). The special education referral and decision-making process for English language learners: Child study team meetings and placement conferences. *The Teachers College Record, 108*(11), 2247-2281.

Liu, K., Albus, D., & Barrera, M. (2011). Moving ELLs with disabilities out of the margins: Strategies for increasing the validity of English language proficiency assessments. In M. Russell (Ed.), *Assessing students in the margins: Challenges, strategies, and techniques.* Charlotte, NC: Information Age Publishing.

Liu, K., Albus, D. & Thurlow, M. (2006). Examining participation and performance as a basis for improving performance. *Journal of Special Education Leadership, 19*(1), 34-42.

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. K., Goldstone, L., Thurlow, M. L., Ward, J., Hatten, J., & Christensen, L. L. (2013). Voices from the field: Making state assessment Decisions for English Language Learners with Disabilities. *National Center on Educational Outcomes*.

National Center on Educational Outcomes. (2011, July). *Understanding subgroups in common state assessments: Special education students and ELLs* (NCEO Brief Number 4). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. www.nceo.info/ OnlinePubs/briefs/brief04/NCEOBrief4.pdf.

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.

Principle 4

Abedi, J. (2004). Accommodations for students with limited English proficiency in the National Assessment of Educational Progress. *Applied Measurement in Education, 17*(4), 371-392.

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

Abedi, J., Zhang, Y., Rowe, S.E. & Lee, H. (2020), Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence-based computer accommodation decision system. Advance online publication. *Educational Measurement: Issues and Practice*. doi:10.1111/emip.12328

Albus, D., & Thurlow, M. (2008). Accommodating students with disabilities on state English language proficiency assessments. *Assessment for Effective Intervention, 33*(3), 156-166.

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

Elliott, J. L., & Thurlow, M. L. (2006). Addressing the needs of IEP/ELLs. *Improving test performance of students with disabilities … on district and state assessments* (2nd ed.). Thousand Oaks, CA: Corwin Press.

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.

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201.

Larson, E. D., Thurlow, M. L., Lazarus, S. S., & Liu, K. K. (2020). Paradigm shifts in states' assessment accessibility policies: Addressing challenges in implementation. *Journal of Disability Policy Studies*, *30*(4), 244-252.

Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A Meta-analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*(3), 10-28.

Shyyan, V., Thurlow, M., Christensen, L., Lazarus, S., Paul, J., & Touchette, B. (2016). *CCSSO accessibility manual: How to select, administer, and evaluate use of accessibility supports for instruction and assessment of all students.* Washington, DC: CCSSO.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*(4), 457-490.

Thurlow, M. L., & Kopriva, R. J. (2015). Advancing accessibility and accommodations in content assessments for students with disabilities and English learners. *Review of Research in Education*, *39*(1), 331-369.

Principle 5

Albus, D. A., Liu, K. K., Thurlow, M. L., & Lazarus, S. S. (2019). *2016-17 publicly reported assessment results for students with disabilities and ELs with disabilities* (NCEO Report 411). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Cortiella, C. (2012). *Parent/family companion guide: Using assessment and accountability to increase performance for students with disabilities as part of district-wide improvement*. Min-

neapolis, MN: University of Minnesota, National Center on Educational Outcomes. (See also movingyournumbers.org)

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.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17*(2), 145-220.

Hauck, M. C., Wolf, M. K., & Mislevy, R. (2016). Creating a next-generation system of K–12 English learner language proficiency assessments. *ETS Research Report Series*, *2016*(1), 1-10.

Telfer, D. M. (2011). *Moving your numbers: five districts share how they used assessment and accountability to increase performance for students with disabilities as part of district-wide improvement*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. (See also movingyournumbers.org)

Thurlow, M. L., Bremer, C., & Albus, D. (2011). *2008-09 publicly reported assessment results for students with disabilities and ELLs with disabilities* (Technical Report 59). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. www.nceo.info/OnlinePubs/Tech59/TechnicalRepport59.pdf.