An Updated State Guide to Universally Designed Assessments

NCEO Report 431

NCEO
National Center on Educational Outcomes

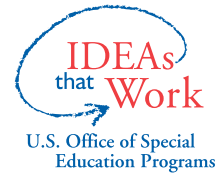# An Updated State Guide to Universally Designed Assessments

Sheryl S. Lazarus, Christopher J. Johnstone, Kristin K. Liu, Martha L. Thurlow, Andrew R. Hinkle, and Kathryn Burden

February 2022

NCEO Core Staff

Executive Summary

This *Guide* is an update to the State Guide to Universally Designed Assessments produced by the National Center on Educational Outcomes (NCEO) in 2006 (Johnstone et al.). It provides a brief overview of what a universally designed assessment is, followed by a set of steps for states to consider when designing and developing, or revising, their summative assessments. The focus of this Guide is states' summative assessments—general content assessments of reading/language arts, mathematics, science, and other content; alternate assessments based on alternate academic achievement standards (AA-AAAS); English language proficiency (ELP) assessments; and alternate-ELP assessments. The purpose of this Guide is to provide states with strategies for designing tests from the very beginning. The objective is to create tests that present an accurate measure of the knowledge and skills of diverse student populations.

This *Guide* describes nine steps designed to improve accessibility of summative assessments:
1. Plan for Universal Design from the Start
2. Define Test Purpose and Approach
3. Require Universal Design in Assessment RFPs
4. Address Universal Design During Item Development
5. Include Universal Design Expertise in Review Teams
6. Perform Usability and Accessibility Testing
7. Implement Item and Test Tryouts
8. Conduct Item- and Test-Level Analyses and Act on Results
9. Monitor Test Implementation and Revise as Needed

Table of Contents ━━━━━━━━━━━━━━━━━━━━━━

Overview

This *Guide* is an update and expansion of one produced by the National Center on Educational Outcomes (NCEO) in 2006 (Johnstone et al.). That guide provided an early set of recommendations for developing universally designed assessments. Much has changed since then, including new technologies that can provide accessibility options for students within the test itself. Although there have been tremendous changes over the past one and one-half decades, the philosophies that guide universal design applied to assessment remain the same.

This *Guide* provides a brief overview of what a universally designed assessment is, followed by a set of steps for states to consider when designing and developing, or revising, their summative assessments. The focus here is states' summative assessments—general content assessments of reading/language arts, mathematics, science, and other content; alternate assessments based on alternate academic achievement standards (AA-AAAS); English language proficiency (ELP) assessments; and alternate-ELP assessments. Although we focus on summative assessments here, we believe the principles apply as well to interim, formative, and diagnostic assessments.

What is Universal Design?

The term **universal design** was originally coined by architect Ron Mace (see Center for Universal Design, n.d.) to describe physical spaces that are amenable to all people who might inhabit them. Mace and colleagues at North Carolina State University outlined seven principles that guide universal architectural design. These are:

- Equitable use

- Flexibility in use

- Simple and intuitive use

- Perceptible information

- Tolerance for error

- Low physical effort

- Size and space for approach and use

Although this *Guide* is not about architecture, the seven universal design principles point to ways in which assessments can be universally designed.

Universal Design for Learning

Universal design is now frequently used in education, for both learning and assessment. The Center for Applied Special Technologies (CAST) is a leader in identifying guidelines for how Universal Design for Learning (UDL) can be used in classrooms. UDL builds on the same principles as universal design for physical spaces. According to CAST, students should have

- Multiple means of engagement

- Multiple means of representation

- Multiple means of action and expression (CAST, 2021)

UDL is based on brain science and the understanding that learners have their individual profiles for how they best engage with, receive, and communicate new learning to others. UDL has also been a useful concept for enhancing inclusion of students with disabilities, English learners, and English learners with disabilities in education systems because it highlights the need to provide students multiple pathways toward learning outcomes.

Universal Design and UDL in Federal Law

The 2015 reauthorization of the Elementary and Secondary Education Act as the *Every Student Succeeds Act* (*ESSA*) directly applies the concept of universal design to assessments. It uses the term "Universal Design for Learning" when doing so. ESSA requires that states use the principles of "universal design for learning" to the extent feasible when developing or revising their assessments that include students with disabilities. The concept of "universal design" also is included in several other federal education laws (see Appendix A).

Universal Design Applied to Assessments

General principles of universal design for physical spaces provide a guide to how universal design is applied to assessments. In 2002, NCEO began considering the importance of assessment accessibility (Thompson et al., 2002). Such accessibility is important because it provides all students with an opportunity to "show what they know and can do" and improves the overall fairness of an assessment for all students, including students with disabilities, English learners, and English learners with disabilities. Because assessment is a different process from architectural design and classroom learning, NCEO provided both a series of underlying principles and a set of elements that describe assessments that are universally designed. The **principles** that undergird universally designed assessments are:

- Universally designed assessments **do not** change the standards measured by assessments. Students may use different pathways to receive and communicate knowledge, but standards remain high for all students.

- Universally designed assessments **should** improve the validity, reliability, and fairness of assessments (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014) for **all students**, including (but not limited to) students with disabilities, English learners, and English learners with disabilities.

Universally designed assessments may **reduce** the need for testing accommodations (National Center on Educational Outcomes, 2021), but there will always be a need for accommodations for some students. Universally designed assessments anticipate the need for one or more accommodations, based on individual student needs and do not make it challenging to use those accommodations..

NCEO's general principles uphold the philosophy that all students have the capability to learn and achieve high standards, yet all students also require assessments that do not present barriers. In this way, the application of universal design to assessment reinforces a systems approach to accessibility and avoids a deficit-based approach. Therefore, applying universal design to assessment requires states and their contracting partners to "test the test" in order to make sure it is accessible to all students. NCEO's early work suggested that designers could evaluate whether elements of universal design were present in assessments. These elements were:

- Inclusive testing population (meaning **all** students)

- Precisely defined constructs

- Accessible, non-biased items

- Amenable to accommodations

- Simple, clear, and intuitive instructions and procedures

- Maximum readability and comprehensibility

- Maximum legibility (Thompson et al., 2002)

NCEO's elements still provide important guidelines for test designers and developers. At the same time, much has changed since NCEO's original conceptualization of universal design. For this reason, this updated *Guide* focuses on expanded approaches to universal design that are useful for states to know when planning for and developing assessments, whether implement-

ing a new assessment or refining current assessments through ongoing item development and test development efforts. We suggest **nine steps** for states to consider as they pursue universal design in their assessments.

Steps for Applying Universal Design to Assessments

This *Guide* provides states with nine steps they can take to demonstrate that universal design is applied to statewide summative assessments. These steps may help states meet federal requirements. More importantly, the incorporation of universal design approaches presents an opportunity for states to improve accessibility for *all* students, including those who have been unable to demonstrate their full potential for achievement because of unintended barriers in assessments.

The steps are designed to improve accessibility of summative assessments. Within each step, there are opportunities for states to consider their own contexts. Further, this *Guide* is not meant to replace or undermine other research currently informing the field of assessment. Rather, the *Guide* is meant to support state assessment leaders as they consider and undertake a series of steps that demonstrate a commitment to universal design principles. The nine steps described in this *Guide* are:

1. Plan for universal design from the start

2. Define test purpose and approach

3. Require universal design in assessment requests for proposals (RFPs)

4. Address universal design during item development

5. Include universal design expertise in review teams

6. Perform usability and accessibility testing

7. Implement item and test tryouts

8. Conduct item- and test-level analyses and act on results

9. Monitor test implementation and revise as needed

Step 1: Plan for Universal Design from the Start

Universal design means thinking about all people who will enter a space, use a product, or interact with a platform. For summative assessments of achievement or English language proficiency, this means planning for an inclusive assessment population.

One of the lessons learned by states and technical assistance providers is that it is critical to know the student populations in their state (Thurlow et al., 2020). Each state has unique demographics, which means that states cannot rely on national data to describe their students. For example, although the percentage of English learners with disabilities in one state is less than one percent, in another state it is more than 20% of the population of school-age students with disabilities. Similarly, the populations of students with disabilities may differ across states in the percentage identified in certain disability categories, just as the primary language backgrounds of English learners vary from state to state.

Planning for universal design from the start requires that assessment designers and consumers understand the characteristics of students in their state. Regardless of past performance, designers and developers should involve knowledgeable stakeholders when defining the purpose and potential use of each assessment. Also, all involved in the process should move away from normative thinking about students, and instead recognize that students have a wide range of abilities, linguistic profiles, and sociocultural backgrounds. Emphasizing the need for high expectations for all students is part of this (Quenemoen & Thurlow, 2019; Thurlow & Quenemoen, 2019). A **commitment** on the part of states, test designers and developers, and stakeholders to making tests accessible to **all** students will support inclusivity that, in turn, will inform other steps in applying universal design to assessments.

Step 2: Define Test Purpose and Approach

State assessment personnel can provide a foundation for universal design in state summative assessments by agreeing on the objectives, constructs, and content that should never go into an assessment (e.g., offensive content or content that reinforces advantages for majority groups), and the desired format of the test and items, including allowable flexibility in presentation and response options for students (CAST, 2021). Allowable accommodations should be identified as well.

Involving stakeholders who know the state's student population well is an important aspect of this step. Knowledgeable stakeholders might include not only assessment experts and content or language development experts, but also advocates for groups of children who typically are excluded from standardized testing practices, including those with disabilities and English learners.

Universal design does not require states to substantially alter many of the technical requirements usually addressed in assessment-related RFPs (e.g., year-to-year equating of tests administered at certain grade levels, establishing performance standards based on specific standard-setting procedures, etc.). Still, a state may want to evaluate the extent to which previous statistical approaches have adequately reflected the diversity of its student population, and suggest or require alternative approaches, as needed.

Step 3: Require Universal Design in Assessment RFPs

Many states seek the assistance of external contractors to undertake work related to test development. Typically, such support is solicited in the form of a request for proposal (RFP) issued to suitable organizations or individuals. Test contractors must then find ways to address the language of the RFP in order to be both responsive and competitive.

States can require evidence of, or a plan for, how bidders will incorporate universal design in assessments by asking for:

- Evidence that the contractor considered the objectives of the assessment

- Evidence that the contractor has considered content that may advantage, disadvantage, or introduce offensive content for any student in the state

- Evidence that the scope of the assessment aligns with state standards

- Evidence that each individual item aligns with a state standard

- A clearly defined construct for each item (which should align with a state standard)

- Identification of target and access skills required to complete each item[1]

- Confirmation that all item writers are knowledgeable about the state's diverse population of students, and are or will be trained in universal design

- Identification of flexibilities that minimize access skill barriers but do not change the construct

- Evidence that allowable accommodations that do not interfere with or change the constructs are incorporated in the assessment

- Evidence that contractors have included elements of universal design, such as clear/intuitive instructions, maximum readability, and maximum legibility

- Examples of items with documentation of how universal design was incorporated in them

In responding to an RFP, bidders must be able to demonstrate how they will develop assessments that are universally designed. The RFP should require that companies that bid on the contract are immersed and have expertise in the principles of universal design. This could also be reflected in RFP requirements related to the qualifications of bidders, reporting requirements, and payment schedules. When applicable, it also should be reflected in the bidders' demonstration of subcon-

---

[1]Ketterlin-Geller (2008) describes the distinction between target and access skills.

tractors' understanding and commitment to universal design, as well as in how the bidder will identify stakeholders to participate throughout the design, development, and revision process.

As states begin to negotiate with their contractors, state assessment personnel can increase the likelihood that their assessments will be universally designed by involving knowledgeable stakeholders to help review and evaluate the types of evidence presented by contractors. These stakeholders, who might include assessment and content experts, as well as advocates for certain groups of children, including those with disabilities and English learners, should know the principles of universal design for assessment.

Step 4: Address Universal Design During Item Development

The development of new items provides a great opportunity to apply universal design, such as when creating a new assessment or adding additional items to an existing assessment. When items are developed from the beginning with accessibility in mind, it may save time and effort later. Well-designed items often move through item-review teams and item tryouts with ease.

Each new item needs to be written with accessibility in mind. This will help to ensure that items respect the diversity of the assessment population, are sensitive to test-taker characteristics and experiences (e.g., gender, age, ethnicity, socioeconomic status, region, disability, language, communication mode), avoid content that might unfairly advantage or disadvantage any student subgroup, and minimize the effects of extraneous factors (e.g., avoid unnecessary use of graphics that cannot be presented in braille, use font size and white space appropriate for clarity and focus, avoid unnecessary linguistic complexity when it is not being assessed). As defined in the test and item specifications, enough items should be developed to cover the full achievement range of the tested population. Item development plans must be sufficient to permit the elimination of items that are not found to meet universal design criteria during pilot testing, field testing, or item analyses.

Item writers often need to be trained on the principles of universal design. Training may include a description of the state's population of test takers and their accessibility needs, as well as how to develop items that are universally designed.

States should:

- Confirm that contractors are well-versed in universal design principles and that they are diverse in representation. A checklist of universal design elements, similar to those provided in Table 1, might be used during item development. It could also be used for state review of items after new items are drafted.

- Ensure item developers are familiar with the state's accessibility and accommodations guidelines and manuals prior to creating any new item.

- For new assessments, ensure that the state's accessibility and accommodations guidelines are developed to be consistent with the targeted purposes, goals, standards, and constructs of the assessment.

- Provide contractors with specifications for the readability and language complexity of items, so that the items reflect the constructs to be tested and do not create access barriers.

Table 1: A Checklist for Item Development

Element	Yes	No	Comments and Recommendations
Does this item appear to consider the entire testing population inclusively? • Does the item include the linguistic, sensory, and experiential diversity of students?			
Are constructs precisely defined? • Does the tested construct align with state standards? • Are there potential access skill barriers in this item?			
Does the item avoid unfairly advantaging or disadvantaging one group?			
If the item has a passage, does the passage avoid unfairly advantaging or disadvantaging one group?			
Are instructions for the item simple, clear, and intuitive?			
Is the item reading level appropriate for the grade level and construct tested?			
If the item has a passage, is the readability and language complexity appropriate for the grade level and construct tested?			
Is all print legible (e.g., 12-14 pt. minimum sans-serif font)?			
Do all graphics support comprehension of the item?			
Do all graphics support comprehension of the item or passage?			
For all graphics, are there ways for students to access the graphic if they are blind or have low vision?			

Table 1: A Checklist for Item Development (continued)

Element	Yes	No	Comments and Recommendations
Are there accessible ways for students to access the item besides print?			
Are there are accessible ways for students to respond to the item beyond standard "clicking"?			

Adapted from Thompson et al., 2002

Step 5: Include Universal Design Expertise in Review Teams

The use of several review teams will help to ensure that an assessment is universally designed. The focus of these review teams might include:

• Accessibility and accommodations policy

• Readability and language complexity

• Item and test content

• Bias and sensitivity

• Section 508 compliance

• Data reviews

Review teams should be well-versed not only in the topic, but also in the principles of universal design, and they should have in-depth knowledge of the student population to be tested. In general, each team should have at least three types of experts: (1) academic content, (2) test taker characteristics and experiences (e.g., cultural, linguistic, specific disability, gender, etc.), and (3) accessibility and accommodations (see Figure 1). Team members should be diverse in representation.

Figure 1: Expertise Needed in Reviews
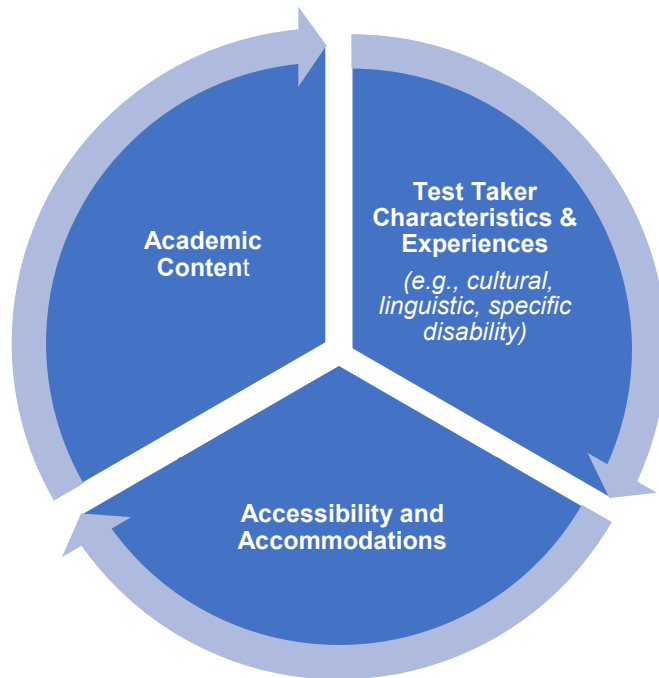
Accessibility and Accommodations Policy

All review teams should include representation from accessibility and accommodations experts. In addition, a separate team devoted to accessibility and accommodations should be formed. This team will oversee not only the policies and training related to accessibility and accommodations, but also provide individuals to participate on other teams, including those, for example, that review language requirements appropriate to the constructs tested; ensure alignment to standards; review potential biases, including experiential or accessibility biases; and conduct 508 compliance reviews.

The Accessibility and Accommodations Team should include, or consult with, content experts because of the connection between tested constructs and allowable flexibility. Tests can be infinitely flexible in terms of presentation, response, and engagement if such flexibility does not change the tested construct. In cases in which the tested construct presents a barrier to a student with a disability, decisions at the state level will need to be made about allowable accommodations.

The Accessibility and Accommodations Team should address all aspects of accessibility. This includes the use of a tiered approach to accessibility (see Figure 2 for an example of how the tiers can be depicted). This approach to setting accessibility and accommodations policy recognizes

that all students have some access needs, that fewer students need individually targeted access features, and also that universal design does not eliminate the need for accommodations, but that many fewer students need them.

Figure 2: Three Level Approach to Accessibility

Universal Features
For all students

Designated Features
For students who need them as identified by an educator in advance

Accommodations
For students with disabilities; in some cases, English learners are also eligible
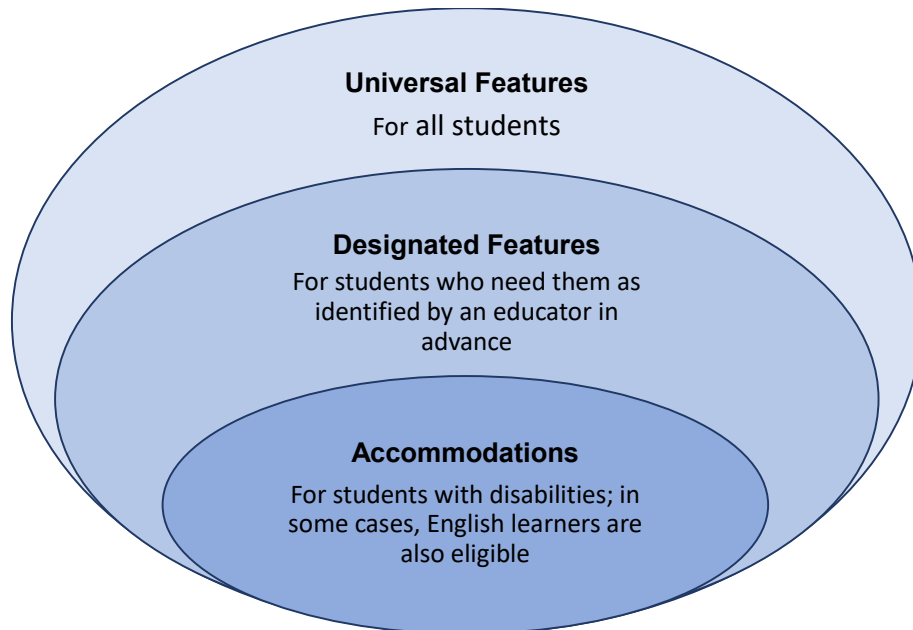
This step confirms that **representation matters.** Nearly every review team, if not all of them, should include accessibility and accommodations experts who can identify potential problems and opportunities for increased access within the frameworks of standards-aligned, bias-reduced items. At the same time, when planning for allowable accommodations, it is important to ensure that intended constructs are not put in danger through flexibility decisions. When built-in accessibility features interact with intended constructs, and students still require access to the test, decisions on accessibility and accommodations policies for specific assessments will need to be made at the state level.

Readability and Language Complexity

The readability and language complexity of reading materials should not exceed grade-level expectations. To ensure assessments are readable and do not introduce access barriers, state personnel should review the readability and language complexity of draft items. Further, reading

and language experts can help to identify areas in which the language of items is unnecessarily complex for the intended construct.

Item and Test Content

A routine part of item development is the evaluation of test items for alignment to state standards (Center for Standards & Assessment Implementation, 2018). If items come from a contractor's item bank, they also must be carefully evaluated for alignment to the state's standards. One way to ensure that aligned items also reflect universal design principles is to include accessibility and accommodations experts on content review teams or, at minimum, teachers of special populations such as teachers of students with disabilities and English learners.

Bias and Sensitivity

Sensitivity review teams examine assessments for items that might introduce bias. Sensitivity teams also examine items for offensive content. These teams may examine patterns in how particular identity groups are portrayed in passages, and examine items for linguistic or experiential biases that may provide an advantage or disadvantage unrelated to tested constructs. Educators and representatives from diverse identity groups play a crucial role in identifying potentially biased items. Sensitivity review teams may recommend striking items entirely from assessments, recommend changes, or support items if they appear to be unbiased. The inclusion of accessibility and accommodation experts in sensitivity review teams will provide further expertise on potential biases that may be introduced in various accessibility settings or under specific accommodated conditions.

Section 508 Compliance

Section 508 of the 1973 Rehabilitation Act requires all government offices to make electronic and information technology accessible to persons with disabilities. Section 508 contains standards and guidelines for electronic communication. These guidelines must be applied to statewide assessments. Some states have passed additional guidance or legislation (U.S. General Services Administration, n.d.). At a minimum, the following aspects should be present in computer-based assessments:

- Assessment allows for font and image resizing by students

- Color contrast is 7:1 for normal text and 4:5:1 for normal font

- Key information is not reliant on color alone for communication

- Redundancy of information is found throughout the assessment

- Text should be provided verbally and textually (unless it interferes with a construct)

- The system should have the capacity to read information for students if appropriate to the construct and the assessment content

- Students can turn off computer reading, change reading speed, control volume

- All stimuli presented in audio AND visual formats

- Alternative text provided for all images

- There are keyboard equivalents for all mouse requirements (Magasi et al., 2017)

Having a Section 508 expert in item review meetings allows for careful examination of the accessibility of items in technology formats.

Data Reviews

A diverse team, including content and bias experts, as well as data analysts with knowledge about accessibility and accommodations, should review items after piloting or field testing. This team should examine performance statistics, including differential item functioning (DIF) of each item (see *Step 8: Conduct Item- and Test-Level Analyses and Act on Results* for more information about DIF analyses) when these data are available. For items that show significant DIF for particular subgroups, reviewers should determine whether it could be an access issue that might be causing DIF for a group of students.

Step 6: Perform Usability and Accessibility Testing

Usability testing should be conducted prior to pilot testing to see how the test features and items function, so adjustments can be made before the test is implemented. As with test reviews, **representation matters.** In usability testing, students can identify potential problems by trying out how items and accessibility features or accommodations function within the administration platform, and providing feedback. Students who represent all populations in the state should be included in usability testing with accessibility and accommodations available to them, as needed.

Additional considerations are needed for English learners and students with disabilities. For example, English learners may have first languages and varying levels of English proficiency. At a minimum, representatives from high and low English proficiency groups should be included in usability studies. Additionally, students with different disabilities should participate in us-

ability studies because a student with one type of disability may not represent the experiences of all students with disabilities. When English learners and students with disabilities participate in usability testing, the objective is to "test the test" by gathering information on the test itself. This means that students should be allowed to provide feedback on the items and test platform functioning in their first language or the language for which they are most proficient (e.g., a language other than English, such as Spanish or American Sign Language) or with any assistive technologies they typically use.

In general, usability testing involves "end users" (i.e., students) working through test items while researchers or usability experts take notes on functionality (U.S. Department of Health and Human Services, n.d.). Usability testing may include cognitive labs, a strategy in which users make verbal or signed utterances explaining what they are thinking about as they proceed through a test (Lazarus et al., 2012). Cognitive labs are an excellent way to identify items that are offensive, confusing, introduce essential non-construct related requirements, or have functionality problems. NCEO has used cognitive labs to capture student experiences with assessments and recommend these labs for any test development process. Procedures may need to be adjusted for students with the most significant cognitive disabilities who participate in AA-AAAS or alternate ELP assessments. Possible adjustments might include having a familiar person administer the items or test, and in some cases, having the adult provide thoughts about how the student is interacting with items.

Step 7: Implement Item and Test Tryouts

Trying out items and tests with larger numbers of students constitutes a pilot or field test that allows for some statistical analyses of how each item functions. The tryout sample should represent the diversity of students found in the state. Some populations (e.g., those with low incidence disabilities) may need to be over-sampled to ensure a large enough sample to derive data for analyses.

An important consideration for this step is that all accessibility features and accommodations are available. To better understand the impact of specific accessibility features or accommodations, cognitive labs can also be conducted during item piloting or field testing, or afterward, to help explain patterns in statistical data.

Step 8: Conduct Item- and Test-Level Analyses and Act on Results

As noted above, when samples are sufficiently large, it is possible to conduct item-level and test-level analyses by subgroup. Descriptive analyses are one way to notice whether populations are scoring differently, but inferential statistics may better help to pinpoint where problems lie with accessibility or bias in items. For such determinations, DIF analyses can determine whether

items function differently for some populations. The same information can be examined for entire tests using Differential Test Functioning (DTF) (AERA, APA, & NCME, 2014).

When items function differently for populations, they should be examined to understand whether differential functioning reflects reasons not related to constructs, such as bias or accessibility barriers. Circling back to sensitivity review teams or cognitive lab data may help identify why items or entire tests function differently by population.

Item Total Correlation (ITC) analyses, such as biserial or point biserial, can also determine whether items function in ways similar to the overall test for individual students or for groups of students (Johnstone et al., 2005). If there are ITC discrepancies by population, there may be bias or barriers in particular items. Careful consideration should be taken to revise or eliminate items that have high DIF or low ITC.

Step 9: Monitor Test Implementation and Revise as Needed

Careful monitoring of test administrations should occur at the local level, in part to ensure that accessibility features and accommodations are administered with fidelity to state policies, Individualized Education Programs (IEPs), 504 plans, and English learner plans if they exist. States can work with districts to ensure that test administration conforms to guidelines and regulations and review data post-administration to identify areas of concern for subsequent years' exams.

States must not simply rely on districts to monitor test implementation. They must have their own plan and means for monitoring test implementation, including adherence to state assessment participation requirements (Hinkle et al., 2021) and accessibility and accommodations policy (Christensen et al., 2009). An important responsibility of the state is to involve its technical advisory committee in reviews of state processes and technical adequacy of each assessment as described in the technical report and other documents. The technical advisory committee must include at least one member with expertise in accessibility. Related to universal design, the technical report should include the entire set of processes used to ensure that the test and its items were developed using universal design principles. It should include, for example, calculations of standard errors of measurement for student subgroups and for accommodated and non-accommodated administrations.

Conclusions

Universal design is not an elusive goal for assessments. Nevertheless, creating an assessment that is universally designed will not be achieved by completing a simple checklist. Rather, creating

assessments that are universally designed requires continuous monitoring and iteration. Critical to realizing universal design are the following steps, as explicated in this *Guide*:

1. Plan for universal design from the start

2. Define test purpose and approach

3. Require universal design in assessment RFPs

4. Address universal design during item development

5. Include universal design expertise in review teams

6. Perform usability and accessibility testing

7. Implement item and test tryouts

8. Conduct item- and test-level analyses and act on results

9. Monitor test implementation and revise as needed

These steps provide a roadmap for improving the accessibility of large-scale assessments. They are not a "one and done" strategy. Rather, they must be considered each testing cycle. Implementing these steps will help to ensure that the assessment will provide all students with an opportunity to show what they know and can do, improving the overall fairness of the assessment for all students, including students with disabilities, English learners, and English learners with disabilities.

Using universal design as an approach when developing an assessment or adding additional items will improve accessibility for a wide variety of users. This will support the creation of tests that produce valid measures of the knowledge and skills of the diverse population of students who take them.

References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing.* AERA.

CAST (2021). *UDL guidelines*. https://udlguidelines.cast.org/

Center for Standards & Assessment Implementation. (2018). *CSAI update: Standards alignment to curriculum and assessment.* CSAI. https://files.eric.ed.gov/fulltext/ED588503.pdf

Center for Universal Design. (n.d.). *About the center: Ronald L. Mace: Last speech, 1998.* https://projects.ncsu.edu/ncsu/design/cud/about_us/usronmacespeech.htm

Christensen, L. L., Thurlow, M. L., & Wang, T. (2009). *Improving accommodations outcomes: Monitoring instructional and assessment accommodations for students with disabilities.* National Center on Educational Outcomes. https://www.cehd.umn.edu/nceo/OnlinePubs/AccommodationsMonitoring.pdf

Hinkle, A. H., Thurlow, M. L., Lazarus, S. S., & Strunk, K. (2021). *State approaches to monitoring assessment participation decisions.* National Center on Educational Outcomes.

Ketterlin-Geller, L. R. (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.

Johnstone, C., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments.* National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/StateGuideUD/UDmanual.pdf

Johnstone, C. J., Thompson, S. J., Moen, R. E., Bolt, S., & Kato, K. (2005). *Analyzing results of large-scale assessments to ensure universal design* (Technical Report 41). National Center on Educational Outcomes.

Lazarus, S. S., Thurlow, M. L., Rieke, R., Halpin, D., & Dillon, T. (2012). *Using cognitive labs to evaluate student experiences with the read aloud accommodation in math* (Technical Report 67). National Center on Educational Outcomes. https://www.cehd.umn.edu/NCEO/OnlinePubs/Tech67/TechnicalReport67.pdf

Magasi, S., Harniss, M., & Heinemann, A. W. (2017). Interdisciplinary approach to the development of accessible computer-administered measurement instruments. *Archives of Physical Medicine and Rehabilitation*, *99*(1), 204–210. https://doi.org/10.1016/j.apmr.2017.06.036

Martinková, P., Drabinová, A., Liaw, Y.-L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE Life Sciences Education, 16*(2), Rm2.

National Center on Educational Outcomes. (2021). *Accommodations toolkit*. https://publications.ici.umn.edu/nceo/accommodations-toolkit/introduction

Quenemoen, R. F., & Thurlow, M. L. (2019). *Students with disabilities in educational policy, practice, and professional judgment: What should we expect?* (NCEO Report 413). National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/NCEOReport413.pdf

Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). National Center on Educational Outcomes. https://www.cehd.umn.edu/NCEO/OnlinePubs/Synth44.pdf

Thurlow, M. L., & Quenemoen, R. F. (2019, May). *Revisiting expectations for students with disabilities* (NCEO Brief #17). National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/NCEOBrief17.pdf

Thurlow, M. L., Warren, S. H., & Chia, M. (2020). *Guidebook to including students with disabilities and English learners in assessments* (NCEO Report 420). National Center on Educational Outcomes. https://nceo.umn.edu/docs/OnlinePubs/NCEOReport420.pdf

U.S. Department of Health and Human Services. (n.d.). *Improving the user experience.* HHS. https://usability.gov

U.S. General Services Administration. (n.d.). *IT accessibility laws and policies: State policy.* GSA. https://www.section508.gov/manage/laws-and-policies/state

Appendix A

Several federal laws refer to universal design. Perhaps the first citation was in the U.S. Assistive Technologies Act (2004). It defined **Universal Design** as—

> a concept or philosophy for designing and delivering products and services that are usable by people with the widest possible range of functional capabilities, which include products and services that are directly accessible (without requiring assistive technologies) and products and services that are interoperable with assistive technology. (Sec 3(19), 2004)

The Universal Design definition aligns with the original conceptualization of Universal Design by the Center for Universal Design (n.d.). It focuses on accessibility for the widest range of people possible without the need for specialized assistive technologies.

Universal Design for Learning was defined in the 2008 Higher Education Act. The framers of the 2008 Act used the above definition for Universal Design in its text, but also added a definition for UDL, which was:

- "A scientifically valid framework for guiding educational practice that – (A) provides flexibility in the ways information is presented, in the ways students respond or demonstrate knowledge and skills, and in the ways students are engaged; and (B) reduces barriers in instruction, provides appropriate accommodations, supports, and challenges, and maintains high achievement expectations for all students, including students with disabilities and students who are limited English proficient" (p. 122, Section 3).

The 2008 Higher Education Act's definition aligns with CAST's framing of UDL. The Act defines UDL as a way to enhance accessibility through flexible means of presentation, response, and engagement in higher education settings.

The Federal **Universal Design for Learning** definition remained the same in the 2015 Every Student Succeeds Act. The Act also represents instances in which Universal Design and UDL must be included in assessment. As part of their reporting requirements for the Act, all States must submit State Plans, may be eligible for State Assessment Grants, and may use grants or internal mechanisms to develop Innovative Assessments. Each of these features (State Plans, State Assessment Grants, and Innovative Assessments) have requirements that States demonstrate Universal Design, UDL, or UDA are planned and administered. Specifically,

- State Plans must outline steps the State has taken to incorporate universal design for learning, to the extent feasible, in alternate assessments" (Section 1111).

- State Grants proposal must demonstrate that States are "developing or improving assessments for children with disabilities, including alternate assessments aligned to alternate academic achievement standards for students with the most significant cognitive disabilities" and "using the principles of universal design for learning" (Section 1201).

- Innovative Assessments must "be accessible to all students, such as by incorporating the principles of universal design for learning" (Section 1204).