



# Interim Assessment Practices for Students with Disabilities

**In spring of 2020 the U.S.**

Department of Education issued waivers to all states, Puerto Rico, and the District of Columbia for administering the required summative assessments in English language arts (ELA), math, and science for the 2019-2020 school year. In the absence of these data many states and districts have been turning to commercially developed interim assessments to get a better understanding of the impact of Covid-19 on student performance and determine the degree to which students are lacking the skills necessary to address grade-level content. Although many of these external assessments can support

educational decision making, they also have the potential to negatively impact individuals and groups of students if not selected and used with caution. This is especially true for students with disabilities<sup>1</sup> who often require specialized support to ensure assessment results provide for valid inferences about the attainment of targeted knowledge and skills.

The purpose of this Brief is to advise the development of guidance that facilitates improved practices related to the use of interim assessments for students with disabilities. It includes a scan of the interim assessment landscape focused on the availability of documentation supporting the appropriateness of these assessments for students with disabilities. The primary sources of information we evaluated for this report included where available: (a) vendor technical reports and manuals, (b) test administration manuals, (c) various documents detailing available accessibility features, and (d) marketing materials. Specifically, these sources of information were reviewed to understand the extent to which commercially available interim assessments were designed to include students with disabilities and the extent to which support is provided for interpreting and using scores.

For the purpose of evaluating specific claims made about the appropriateness of specific interim assessment uses, we reviewed these materials with a set of organizing questions. Do vendors:

- explicitly or implicitly identify students with disabilities as part of the targeted test population?
- provide alternate assessments for students with the most significant cognitive disabilities?
- provide evidence of detailed attention to the principles of universal design and involvement of experts in special education and students with disabilities during test design, development, and standard setting?
- make accessibility features available to students with disabilities?

---

<sup>1</sup>Students with disabilities include students who have an Individualized Education Program (IEP) and those who have a 504 accommodations plan.

We also reviewed the sources of information with questions about the appropriateness of score interpretations for students with disabilities, including:

- When students with disabilities are included in the target population (explicitly or implicitly), is there evidence of the appropriateness of their inclusion?
  - Beyond alignment evidence presented overall for all students, is there specific evidence that alignment was examined between the supported interpretations and the intended uses for students with disabilities?
  - Is there evidence of measurement invariance between students with disabilities and their peers without disabilities?
- Are the intended purposes and uses explicitly supported for students with disabilities?

Although a broad range of commercial interim assessments was reviewed, given the large number of products on the market this review was not exhaustive. Instead, we identified a collection of commonly used products that varied with respect to their intended purpose and design and for which at least some technical documentation was publicly available. Our selection of interim assessments represented products developed by ACT, Curriculum Associates, Fountas and Pinnell, NWEA, Pearson, Renaissance Learning, Smarter Balanced, and University of Oregon's Center on Teaching and Learning.

It is important to acknowledge from the onset that documentation was reviewed with the goal of establishing a broad understanding of the type and range of information available to support decisions about assessment use and quality for students with disabilities. The absence of evidence does not indicate that it does not exist, only that it was not referenced or located in our review of publicly available documentation. The focus on publicly available documentation serves to inform guidance by reflecting on the transparency of technical information available to inform test selection,

evaluation, and use. However, discussions with interim assessment vendors may further clarify how and when validity evidence focused on students with disabilities is collected and reported to stakeholders for consideration.

It is also important to note that this report reflects on the quality and scope of evidence supporting the intended purposes and uses outlined by assessment vendors. Because the locus of control for these types of assessments is typically the district or school, additional research is needed to understand whether and how local uses go beyond those suggested and validated by test vendors. Although state education agencies may play a role in supporting or promoting local implementation, decisions about test administration and use are often made at a local level. In the absence of clear guidance and oversight from state departments of education—of the type that typically accompanies large-scale state summative assessments—districts, schools, and educators may use these tests in ways that are not supported, especially in the current context where the demand for information is high and access to high quality test data is scarce.

This Brief is structured in three sections. Section 1 reflects on a common definition of interim assessment and highlights the manner and degree to which the large array of assessments sharing this label can differ. Section 2 addresses the inclusion of students with disabilities in the intended test taking population, lists common intended uses of interim assessments, and summarizes the manner and degree to which interim assessment documentation supports the appropriateness and utility of these assessments for students with disabilities. Section 3 outlines additional factors that should be considered when determining how best to help states support local efforts to check the use of interim assessments for students with disabilities.

## Section 1: Interim Assessments

In 2009 Perie et al. (2009) provided the following definition of interim assessments:

Assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purpose and intended uses, but the results of any interim assessment must be aggregable for reporting across students, occasions, or concepts. (p. 6)

The generality of this definition reflects the diverse range of products currently referred to as interim assessments. They are, essentially, any tools that can be used to inform teaching and learning throughout a course of instruction. Because they address the information gap between summative and formative assessment, interim assessments vary significantly in function and design. They may be designed to measure a broad range of content that serves to predict students' performance on a state's summative assessment at fixed points throughout the school year, or inform teachers' formative assessment strategies by evaluating students' understanding of one or more skills needed for success in an upcoming unit of instruction. Despite this diversity, tests sharing the interim assessment label often are referenced as if they are interchangeable and marketed in ways that suggest the same test can support multiple, often competing goals equally well. Furthermore, the benefits and shortcomings of interim assessments often are discussed using generalities that can interfere with state and district leaders' efforts to critically evaluate these assessments for their specific information needs. For these reasons, some have suggested that it would be more productive if these assessments were referenced and distinguished in terms of how they are used rather than with the common "interim" label (D'Brot & Landl, 2019).

### Dimensions of Variation in Interim Assessments

Figure 1 uses common characteristics of summative and formative assessment to represent the ends of a hypothetical interim assessment continuum that varies along multiple dimensions. As shown,

summative assessments are tests administered at the end of a grade or course typically for accountability or program evaluation purposes. They are designed to prioritize score reliability and comparability and support inferences about student performance against end-of-grade or course expectations. In contrast, formative assessment is an ongoing process that educators engage in during instruction to collect evidence of student learning. The information gained is used by teachers to adjust instruction and by students to evaluate and monitor understanding of targeted concepts and skills.

A key dimension of the variation in interim assessments is the grain-size of the target of measurement. For a given test, the target of measurement is the set of knowledge, skills, and understandings that must be measured in order to interpret the results in a manner that supports the

intended use of results. For example, in order to use the results of an assessment to *monitor students' progress toward end of year expectations in Grade 7 math* the assessment must be designed to produce a score that can be interpreted as reflecting a students' current understanding of the expected grade 7 math concepts and skills. Although additional design features are necessary to evaluate progress over time, the target of measurement is the set of knowledge and skills that support this inference.

Based on our review, interim assessments can be classified in one of four levels reflecting differences in the granularity of the target of measurement. These levels and a description of each are provided in Table 1.

If appropriately designed, assessments at any of these levels may be used to:

**Figure 1. Continuum of Assessment Design Features and Uses**

Dimensions of Variation		Summative Assessment	Interim	Formative Assessment
<b>Design Features</b>	<b>Target of Measurement</b>	<ul style="list-style-type: none"> <li>Entire course or grade level expectations</li> </ul>	↔	<ul style="list-style-type: none"> <li>The skill/standard that is the focus of instruction</li> </ul>
	<b>Properties of Results</b>	<ul style="list-style-type: none"> <li>Comparable</li> <li>Reliable</li> <li>Generalizable</li> <li>Accurate</li> </ul>	↔	<ul style="list-style-type: none"> <li>Granular</li> <li>Instructionally useful</li> <li>Provides insight into student thinking</li> </ul>
	<b>Time of Administration</b>	<ul style="list-style-type: none"> <li>At the end of a course of instruction</li> </ul>	↔	<ul style="list-style-type: none"> <li>During teaching and learning</li> </ul>
	<b>Administration Conditions</b>	<ul style="list-style-type: none"> <li>Standardized</li> <li>Secure</li> <li>Scheduled</li> </ul>	↔	<ul style="list-style-type: none"> <li>Frequent</li> <li>Informal</li> <li>On-demand/flexible</li> </ul>
	<b>Goal of Score Reports</b>	<ul style="list-style-type: none"> <li>Summarize</li> <li>Categorize</li> <li>Classify</li> <li>Rate</li> </ul>	↔	<ul style="list-style-type: none"> <li>Describe</li> <li>Diagnose</li> <li>Prescribe</li> </ul>
<b>Primary Uses</b>		<ul style="list-style-type: none"> <li>Accountability</li> <li>Report achievement status</li> <li>Report growth</li> <li>Program evaluation</li> <li>Inform resource distribution</li> </ul>	↔	<ul style="list-style-type: none"> <li>Inform teaching and learning process.</li> <li>Adjust instruction in the moment</li> </ul>

**Table 1. Levels of the Target of Measurement**

	Description
Level 1. Summative Domain	Sample content from the entire domain associated with a grade or course such as English language arts (ELA), math or science. Often referred to as mini-summative assessments because they represent the range and complexity of content measured on the end-of-year summative exam and may report out on the same or similar reporting categories.
Level 2. Sub-Domain	Provide information about student performance in a large sub-domain of a content area, such as reading or writing.
Level 3. Reporting Category	Provide information about student performance on a set of related skills or standards such as those associated with a defined reportable category on the state summative exam, an important learning goal or a big idea of the discipline.
Level 4. Focal Skills/Standards	Designed to measure student performance on a narrow set of skills or standards.

- understand current achievement,
- monitor within-year progress,
- evaluate the impact of instruction on performance, or
- identify professional development needs at the district, school, or teacher level within the targeted domain.

What differs across levels is the degree to which the assessment results provide information that directly informs instruction and provides students with individualized feedback and targeted supports. The more focused the target of measurement, the more useful the results will be in helping students and teachers understand the actions necessary to change performance (Marion, 2019).

Because stakeholders need different types of information to support decision making, many vendors offer multiple interim assessment products spanning the levels represented in Table 1. Although these “assessment systems” provide stakeholders a broader array of tools to collect information, they also increase the likelihood of misuse and over-reliance on test data in the absence of appropriately targeted professional development.

### Score Interpretation

Scores that inform broad claims about students’

level of achievement can support educators by differentiating performance in meaningful ways and shining a spotlight on struggling students. Consequently, achievement levels and corresponding descriptions are common features of most interim assessments. Typical determinations made with information included on interim assessment reports include a student’s:

- Proficiency or benchmark level
- Mastery
- Growth
- On track designation
- On grade designation
- Readiness
- Risk status

In order to use the results in these ways for students with disabilities, evidence must be provided that the scores mean the same thing for these students and do not result in unintended negative consequences.

## Section 2: Summary of Evidence

Our review of vendor information focused on the identification of evidence that scores have the same meaning for students with disabilities as for other

students and do not result in unintended negative consequences. This section summarizes the nature and level of validity evidence that we found supporting claims (explicit or implied) that the intended uses of test scores are appropriate for students with disabilities. Identified gaps in evidence can highlight areas of concern or limitations of these assessments to fully serve the needs students with disabilities, and in turn inform the development of guidance to support improved stakeholder evaluation and use.

Our findings are drawn from publicly available documentation for a selection of 13 commonly used interim assessments produced by the eight test vendors. In this review, we relied primarily on administration and accessibility guides, technical reports and manuals, and various marketing materials and statements available on test vendor sites, including white papers.

For the 13 tests that were reviewed, technical manuals and accessibility guidance documents were available directly on the vendors' website for four tests. Technical documentation was available by request for six more of the 13. We were unable to identify any technical documentation or specific information about the availability of accessibility features for students with disabilities for the remaining three interim assessments in our selection.

### **Inclusiveness**

Both the Every Student Succeeds Act (ESSA, 2015), and the Individuals with Disabilities Education Act (IDEA, 2004) call for the inclusion of students with disabilities in assessments. They indicate that most students with disabilities will participate in general assessments, with accommodations as needed. A small percentage of students will participate in alternate assessments based on alternate academic achievement standards for students with the most significant cognitive disabilities. An alternate assessment is to be developed and implemented for each state and districtwide assessment.

To determine the level of inclusiveness for students with disabilities in our review, we evaluated

documentation of several accessibility-related factors. We looked for evidence that students with disabilities can be assessed under conditions that support their specific needs. Specifically, we looked for evidence of the availability of the following:

1. Universal design
2. Designated supports
3. Accommodations
4. Special forms
5. Alternate assessments

Our review focused only on how vendors support the accessibility needs of students with disabilities. We did not examine the extent to which vendors attempted to support the important processes of identifying accessibility needs of individual students with disabilities or monitoring the implementation and use of accessibility features. Support for these processes (e.g., specific guidance on the importance of appropriate identification of student needs for and use of supports in the assessment process) would provide end-to-end support for students with disabilities from identification through interpretation and use of scores.

We noted that vendors with both summative and interim assessments tended to have clear and more comprehensive documentation of the accessibility features offered. We speculate that peer review requirements (U.S. Department of Education, 2018) have played a large role in organizational thinking, planning, and development to meet the accessibility needs of students with disabilities, with systems and protocols built specifically to support the needs of these students. Where vendors provide more than one type of assessment (e.g., spanning the levels referenced in Table 1), we noted that the same accessibility features tend to be uniformly available over those assessments.

### **Universal Design**

The Higher Education Opportunity Act (2008) defines universal design for learning as:

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 [English learners].

The National Center on Educational Outcomes (NCEO) (Thompson et al., 2002) identified seven universal design elements that are specific to assessment:

- **Inclusive assessment population:** all students have the opportunity to participate in the assessment
- **Precisely defined constructs:** construct-irrelevant variance is mitigated for all students
- **Accessible, non-biased items:** content does not advantage or disadvantage any groups
- **Amenable to accommodations:** design features facilitate the use of accommodations
- **Simple, clear, and intuitive instructions and procedures:** language is used that supports student understanding of what they are being asked to do
- **Maximum readability and comprehensibility:** probability of comprehension by different groups of students is determined
- **Maximum legibility:** legibility of all content is demonstrated: text, graphs, tables, and illustrations

The practice of following universal design procedures during content development and test construction provides an important means for students with disabilities and English learners to access the intended construct without first having to decipher non-construct relevant material or features that may be present in a test's content. For example, removing or limiting overly complex, or unnecessary language is one way that serves the needs of all students, but has a particular positive effect in allowing English

learners and students with specific language-related disabilities to avoid interference of language that is irrelevant to the knowledge, skills, or abilities a student is expected to demonstrate. Universal design is a solution that balances accessibility needs with standardized administration procedures.

Vendors for eight of the 13 interim assessments reviewed provided at least some information about the universal design principles that were followed during test development and content reviews. Materials for the remaining five assessments were silent on the matter of universal design. There was a fair amount of variation in the comprehensiveness of the universal design discussions, ranging from providing detailed rationales and evidence that universal design principles are routinely followed, to minimal references to the use of universal design principles during content development and test construction. We found only one interim assessment vendor that specifically referenced the use of NCEO universal design principles (Thompson et al., 2002) for each of the three interim assessments it offers.

### Designated Supports

Designated supports are features that can be used by any student for whom the need has been determined by an educator or team of school-level decision makers. The fundamental difference between designated supports and accommodations is that the latter are typically determined by a formal Individualized Education Program (IEP) or 504 accommodations planning team. This means also that there tends to be some overlap in how designated supports and accommodations are classified.

For example, small group settings are sometimes classified as accommodations, and sometimes as a designated support.

Designated supports are not expected to change the measurement properties of the test or present challenges to score meaning for students using them. Examples include:

- Testing individually or in a small group
- Access to food, drink, medications during testing

- Using colored overlays for paper testing
- Magnification of the test content

Of the 10 out of 13 interim assessments for which accessibility information was available in this review, vendor documentation of available designated supports for their interim assessments ranged from comprehensive lists and procedures for score interpretation and use, to an absence of any reference to such supports. Documentation for one test provided detailed information about the availability of, and procedures for the use of designated supports. Documentation for seven of the 10 provided relatively complete lists of available designated supports, but little or no information guiding their use. It is expected that specific guidance for the use of designated supports would be included in test administration manuals provided to test users after purchase. Available documentation for two of the tests provided no information about the availability of designated supports.

### Accommodations

The Americans with Disabilities Act (ADA, 1990) defines testing accommodations as, “...changes to the regular testing environment and auxiliary aids and services that allow individuals with disabilities to demonstrate their true aptitude or achievement level on standardized exams or other high-stakes tests.” The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education’s (NCME) *Standards for Educational and Psychological Testing* (2014) define accommodations similarly but tailor its version to the conception of accommodations as the means to provide access to the construct in way that does not change the meaning of examinee scores. The *Standards* state “Accommodations consist of relatively minor changes to the presentation and/or format of the test, test administration or response procedures that maintain the original construct and result in scores comparable to those on the original test” (p. 58). Where the ADA emphasizes the requirement to appropriately lower barriers for individuals with disabilities so that they may fully

demonstrate their abilities, the *Standards* also place a heavy emphasis on score comparability.

Policies and guidance define the accommodations available for state tests. IDEA requires that states report the numbers of students with IEPs who are provided accommodations. Vendors generally produce these reports and include more detailed information (e.g., for specific accommodations) in technical manuals. Accommodations often are classified into four categories:

1. **Timing and Scheduling:** allows flexibility for how the test time is organized. For example, a student who requires extra time to take an assessment may need multiple sittings to complete longer tests.
2. **Presentation:** reduces barriers in access to the test content. For example, a read-aloud may be provided for students with specific language disabilities, or to students with impaired vision.
3. **Setting:** allows changes in the location or conditions of the testing place. For example, a student may be tested in an individual or small group setting to mitigate the effect of distractions.
4. **Response:** reduces disability-related barriers to a student demonstrating the requisite knowledge, skills, and abilities by allowing them to complete assessment tasks in different ways. For example, a student that is unable to physically write a response may use a scribe, or speech-to-text tool.

Evidence of the availability of accommodations for interim assessments was not uniformly comprehensive. A few vendors provide details on the suite of accommodations offered under each category, but many of the interim assessments that were reviewed here provided little or no evidence of the availability of a wide range of accommodations. For some interim assessments, no accommodations appear to be readily available.

Of the 10 tests for which accessibility information was available for this review, vendor documentation



of available accommodations for their interim assessments ranged from comprehensive lists and procedures for score interpretation and use, to nominal reference to one or two accommodations. Documentation for five of the 10 tests provided quite detailed information about the availability of, and procedures for the use of accommodations. Documentation for five of the tests provided either no information about the availability of designated supports or simple references to one or two accommodations (e.g. an audio option for visually impaired students).

### **Special Forms**

IDEA requires that the same rigorous expectations are maintained for students with disabilities as for the general population of students, accompanied by appropriate accessibility support to allow students to fully demonstrate their knowledge, skills, and abilities against those expectations. Special forms are intended to assess the same content as for students who do not have special accessibility needs. Examples of special forms include braille, large print, and translated test forms.

Our review of interim assessment documentation showed that seven of the 13 interim assessments reviewed are available in Spanish. We also found that seven of the 13 reviewed interim assessments are available in braille, either in paper or refreshable braille.

### **Alternate Assessments**

Alternate assessments were first required by IDEA in 1997 (Sec. 1412(a)(16)). As part of the requirements for state IDEA funding, the state (for the state assessment) or the district (for a districtwide assessment) must develop and implement “guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in regular assessment... with accommodations ...in their IEPs.” Subsequently, ESEA and IDEA amendments and regulations clarified that alternate assessments based on alternate academic achievement standards are for students with the most significant cognitive disabilities.

Our review of the interim assessment landscape did not identify any alternate interim assessments.

### **Score Interpretation and Use**

A challenge for evaluating the appropriateness of the intended score interpretations and uses for students with disabilities is that not all vendors provide evidence in their technical documentation that general claims are met for students with disabilities. Several vendors make explicit claims that the intended purposes and uses are valid for *all* students, implying inclusion of students with disabilities in the established validity basis for test score interpretations.

The *Standards* detail the five sources of validity evidence as evidence based on (a) test content, (b) response processes, (c) internal structure, (d) relations to other variables, and (e) the consequences of testing. Typical evidence in a summative assessment context includes detailed discussions of test specifications and design details, content reviews for relevancy, sensitivity, and bias, cognitive laboratories, and detailed technical and psychometric results (e.g., group reliabilities, classification accuracy and consistency, model and person fit, differential item functioning, dimensionality, correlation of scores with measures of similar content, etc.). The contrast between evidence provided in support of the interim assessments reviewed here, and state summative assessment more generally was conspicuous, even for the general population of students.

The validity evidence provided by vendors for the 13 interim assessments reviewed show a range in the comprehensiveness and quality of support for the validity of intended score interpretations. Evidence supporting use claims for all students ranged from reasonably complete attention to each of the five sources of validity described in the *Standards*, to no supporting evidence that scores are reliable and valid for their intended uses.

Evidence of the validity of score interpretations for students with disabilities was largely absent, even for the most well documented validity arguments

for the general population. Students with disabilities were identified as focal groups in differential item functioning (DIF) and standard error of measurement statistics in one suite of three assessments offered by a single vendor. In one other suite of three assessments, reliabilities were provided separately for students with disabilities. This dearth of information on the validity of scores for students with disabilities demonstrates a pervasive lack of evidence that interim assessment scores for the assessments reviewed can be interpreted similarly to the general population of students.

Two vendors did make claims specific to students with disabilities by way of stating that the assessment is appropriate for use in screening for dyslexia. However, no technical support for that claim was found in the publicly available documentation. Also, validity evidence in support of scores based on special forms (Spanish and braille formats) is absent.

### **Summary of Gaps in Documentation**

The following provides a summary of the specific gaps between claims (explicit or implied) for students with disabilities and evidence to support them that were noted based on the documents reviewed for the 13 assessments in our selection.

**Marketing Materials.** Most marketing and technical documentation either directly refers to, or indirectly implies that *all* students are included in the intended population. Guidance that clarifies the conditions that must hold in order for scores to be interpreted and used as intended and considers the needs of specific student populations would help test users understand the limitations of score use for individuals and groups of students, including students with disabilities.

**Statistical Evidence of Measurement Invariance.** In general, statistical evidence that scores for students with disabilities have the same meaning as scores for other students is lacking. In comparison, reliabilities, classification accuracy/consistency, model and person fit, DIF, dimensionality (e.g., confirmatory factor analysis, weighted multidimensional scaling, and principle components analysis with a parallel analysis),

and the results of other test property invariance analyses are routinely reported in large-scale summative assessment technical documentation.

**Growth.** No documentation was found that provides evidence that growth measures (regardless of the metric used) have the same meaning for students with disabilities.

**Construct Definition.** There is a range of rigor in which universal design principles are applied to content and test development processes and procedures. Most vendors articulate the content to be measured at least generally, but few provide technical documentation that is explicit about the connection between construct definition and how item design specifically avoids distracting or extraneous (construct irrelevant) factors associated with the principles of universal design. To the extent that such procedures isolate the most important elements of the construct, accessibility features are less likely to interfere with students' ability to demonstrate their standing fairly. This is particularly useful given the limitations associated with small sample sizes that are typically available for analysis. Nor does the technical documentation routinely provide details about the involvement of experts in students with disabilities in the development process.

**Response Process Validity.** No evidence was found that protocol analyses were used in cognitive laboratory-style studies to support an understanding of any elements of the test content and presentation that are challenging for students with disabilities in particular. Small sample analyses would be useful to evaluate whether supports, accommodations, and special forms are actually helping students with disabilities to access the construct.

## **Section 3: Considerations Informing the Development of Guidance**

Our review highlighted several factors that need to be considered when establishing guidance to help state and district leaders make informed decisions about interim assessment use—in general and specific to students with disabilities. These include, but are not

limited to, the role of the state in supporting local implementation of interim assessments, expectations related to the availability of validity research and data, and the need for greater clarity about local uses of assessment results for students with disabilities. Each of these factors is discussed briefly.

### **1. Role of the State in Supporting Implementation of Interim Assessments**

A scan of state Department of Education websites shows that there are several ways in which states may support or promote the use of interim assessments at a local level. For example a state may:

- mandate the administration of a state selected or developed interim assessment (e.g., Arkansas requires all districts to administer one of four state-procured interim assessments in K-2 for math and ELA).
- offer one or more state-purchased interim assessment tool for use by districts on a voluntary basis (e.g., Oregon offers the Smarter Balanced Interim Assessments and Tools for Teachers; Pennsylvania offers its state-developed Classroom Diagnostic Tools).
- identify a set of approved or endorsed interim assessments or providers (e.g., District of Columbia).
- provide general guidance and professional development that informs local assessment evaluation and selection efforts (e.g., Rhode Island's Guidance for Developing and Selecting Quality Assessments in the Secondary Classroom).
- take no role in informing decisions about interim assessments at a local level.

The role states play in supporting local assessment initiatives is likely to influence the impact they have on local interim assessment use practices (in general and for students with disabilities). Different strategies may be necessary for states that are more or less involved in local efforts to design or implement assessments other than the

state summative assessment.

### **2. Transparency of Data and Research**

In many cases access to expected validity data was limited, difficult to find, or not publicly accessible. In some cases, research could be found, but only through a comprehensive internet search and then it was associated with a previous version of the assessment. Guidance should serve to empower those selecting and using interim assessments by helping them understand, identify, and request reasonable evidence of technical quality, in general and for students with disabilities. It should clarify not only the type of evidence necessary to evaluate the degree to which assessments support students with disabilities, but also the frequency with which that evidence should be reported and updated to guarantee the vendor is doing its due diligence.

### **3. Availability of Validity Data**

Although vendors provide general guidance to support administration of their assessments, decisions are typically made at a local level to support district or school goals. Even for interim assessments that are administered statewide within a specified testing window, administration conditions and the collection of student-level demographics is likely to vary across districts. Consequently, there is often limited empirical validity evidence supporting the proposed interpretation and use of these assessments, in general and for specific student groups. If demographic data are collected, N-counts for student groups may be small, or those data may not be made available to vendors. For this reason, special studies often are necessary to collect trustworthy information about the appropriateness of the assessment for student groups. Guidance should explain why empirical validity evidence may be lacking for some student groups and establish criteria that support local decisions about the adequacy of evidence provided relative to the intended use of results.

#### 4. Understanding Local Score Use

Although we can identify the proposed uses of interim assessments by vendors, the specific ways in which they are being used by districts and schools (especially for students with disabilities) is unknown. We can anticipate that they may be using them to support decisions for which they are not intended (e.g., identifying and tracking IEP goals) but we will not know without additional research. Surveys should be administered to better understand the ways in which interim assessments are used by stakeholders to inform decisions about students with disabilities. In this way, tools developed to prepare district and school leaders to evaluate and discuss the appropriateness of interim assessments for students with disabilities can address all of the ways in which these assessments are currently being used.

#### 5. The Need for Curricular Specificity

The validity and instructional utility of off-the-shelf interim assessment results are threatened if the assessment does not reflect the learning objectives and strategies reflected in curriculum and instruction. Such validity issues may be compounded in the case of students with disabilities for which specific instructional techniques or learning trajectories may be defined to support the attainment of individual goals. Guidance should highlight the importance of evaluating interim assessments in terms of coherence with curriculum and instruction for students with disabilities in addition to required accessibility features.

#### 6. Absence of Alternate Interim Assessments

Our analysis did not identify vendors that offered alternate interim assessment. Guidance developed to inform states should clarify this point, so it is clear that there is currently no tool that supports the inclusion of these students in an interim assessment administration. This lack of alternate interims is important if an interim is suggested as a way to meet accountability requirements or to support other uses that require a large-scale census administration that includes all students.

## References

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

Americans with Disabilities Act, 104 U.S.C. § 328 (1990).

D'Brot, J, & Landl, E. (2019, June 11). Stuck in the middle with interim assessments: Improving the selection, use, and evaluation of interim assessments. *Centerline*. <https://www.nciea.org/blog/assessment/stuck-middle-interim-assessments>

Every Student Succeeds Act, 114 U.S.C. § 95 (2015).

Higher Education Opportunity Act, PL 110-315 (2008).

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Marion S. (2019, March 21). Five essential features of assessment for learning. *Centerline*. <https://www.nciea.org/blog/innovative-assessment/five-essential-features-assessment-learning>

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5-13.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). National Center on Educational Outcomes. <http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis44.html>

U.S. Department of Education. (2018). *A state's guide to the U.S. Department of Education's assessment peer review process*. <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>

## NCEO Brief #22, April 2021

This Brief was written by Michelle Boyer and Erika Landl. It was published jointly by the National Center on Educational Outcomes (NCEO) and the National Center for the Improvement of Educational Assessment.

NCEO Director, Sheryl Lazarus; NCEO Assistant Director, Kristin Liu.

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Boyer, M., & Landl, E. (2021, April). *Interim assessment practices for students with disabilities* (NCEO Brief #22). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes and National Center for the Improvement of Educational Assessment.

NCEO is supported through Cooperative Agreements (#H326G160001) with the U.S. Department of Education, Office of Special Education Programs, Research to Practice Division. The Center is affiliated with the Institute on Community Integration at the College of Education and Human Development, University of Minnesota. Project Officer: David Egnor.

Opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education.

This publication is available in alternative formats upon request. Direct requests to:

National Center on Educational Outcomes  
University of Minnesota • 207 Pattee Hall  
150 Pillsbury Dr. SE • Minneapolis, MN 55455

Phone 612/626-1530 • Fax 612/624-0879

*Visit our website at [www.nceo.info](http://www.nceo.info)*

INSTITUTE on COMMUNITY INTEGRATION

UNIVERSITY OF MINNESOTA

The University of Minnesota is an equal opportunity employer and educator.