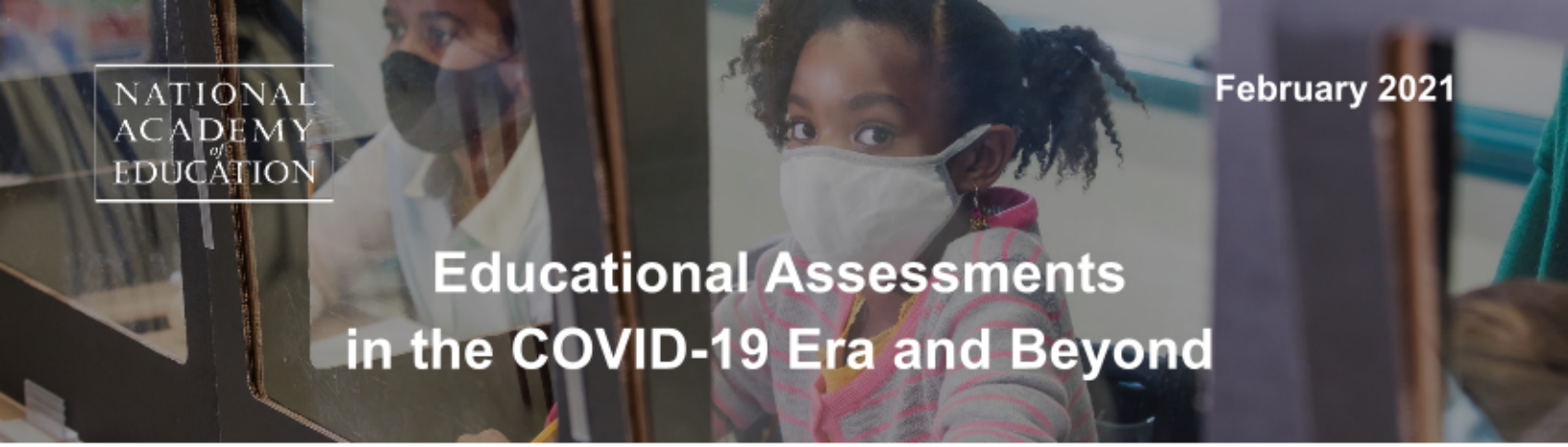

NATIONAL ACADEMY of EDUCATION

February 2021

Educational Assessments in the COVID-19 Era and Beyond

[Concerning] the children of this pandemic … [t]he models no longer apply, the benchmarks are no longer valid, the trend analyses have been interrupted…. When the children return to school, they will have returned with a new history that we will need to help them identify and make sense of…. There is no assessment that applies to who they are or what they have learned.

Dr. Teresa Thayer Snyder,
Retired Superintendent, Voorheesville Central School District, NY[1]

[G]iven a shortage of testing data for Black, Hispanic and poor children, it could well be that these groups have fared worse in the pandemic than their white or more affluent peers…. Given these realities, the new education secretary … should resist calls to put off annual student testing.

Editorial Board,
The New York Times[2]

INTRODUCTION

As the COVID-19 pandemic continues to spread across the United States and around the world, school systems everywhere are in crisis management, with education leaders and teachers struggling to provide continuous instruction via combinations of in-person, virtual, and hybrid learning modes. In this uncertain and fluid environment, the regular challenges of assessing what and how students are learning have become even more complex: teachers need information to guide classroom-level learning—no matter which instructional mode—and states, school districts, schools, teachers, parents and caregivers, students, and communities need evidence of how COVID-19 is affecting historically marginalized, disadvantaged, and underserved students.

The two quotes above reflect diverse opinions about what information regarding student learning is most needed, the critical audiences for that information, and the most appropriate ways to obtain it in the remainder of the current school year and for the next school year beginning in fall 2021. Although reliable data are necessary to inform future educational goals and resource allocations, how these data are gathered, and ultimately used, is contested. The persistent debate about fair uses of assessment for instructional improvement and accountability has become more heated, as educators and policy makers weigh the benefits and risks of suspending mandatory assessment requirements under the federal Every Student Succeeds Act of 2015 (ESSA), which requires annual summative assessments in grades 3–8 and once in grades 10–12. At the crux of this argument is the balance between the fairness of holding schools, teachers, and students

accountable for performance under the arduous conditions imposed by the pandemic and the equally compelling logic behind maintaining a flow of valid information on whether (and which) students are learning and in which contexts.

A fundamental and familiar question, therefore, centers on the *rationale for assessment*. What are its goals and, in particular, can assessment advance teaching and learning and reduce educational inequities? In the near term, what are the best "uses" of assessment in 2021?

To address these questions, the National Academy of Education (NAEd) convened a group of scholars, policy leaders, and educators (see the list attached to this summary report) for a focused discussion of the "how" and "why" of testing in both the contexts of the special circumstances of 2021 and beyond. This online roundtable built on NAEd's prior work addressing COVID-19 as well as its historical[3] and recent[4] work addressing educational assessments. Presented here are some of the overarching themes of the conversation to stimulate further discussion among educators, researchers, policy makers, and the general public.

This summary report begins where the roundtable conversation kept returning, with a description of the purposes and intended users of different types of assessments. Next, it discusses inequities in education and implications for the appropriate uses of assessment. Then the report addresses the 2021 end-of-year "summative" assessments: assuming that school districts will administer such assessments, it points out caveats to keep in mind regarding test administration, interpretation, and intended and unintended uses of test results. Finally, looking beyond 2020–2021 end-of-year assessments, the report discusses themes that emerged, including ensuring that assessment systems are balanced and equitable, reframing accountability from a deficit lens to an improvement perspective, and expanding assessment literacy.

PURPOSES, AUDIENCES, AND TYPES OF ASSESSMENTS

Educational assessment[5] is a process for obtaining information that can be used for making decisions about students; teachers, curricula, programs, and schools; funding; and other aspects of educational policy. There are numerous audiences and users of information obtained from assessments. For example, parents and caregivers may use test scores to understand how their children's opportunities and achievement compare to other students in the class or school. Teachers may use test scores to determine areas to focus additional and varied instruction. Schools, districts, and states may use test scores to monitor student performance on a more macro level, document and highlight inequities in the system, make graduation and placement decisions, allocate funds, evaluate teachers, and determine professional development needs. The federal government mandates assessments as part of an accountability system to ensure equal educational opportunities for all children. Such accountability takes many forms, including estimates of academic growth and trends over time. Other uses of assessment relate to decisions outside the realm of instruction and curriculum (e.g., home buyers who include test scores at the school and local level to inform purchasing decisions). [6]

No single test can serve all of these purposes with requisite validity and reliability.[7] Critically, the intended purposes and uses of a test should be defined and explicitly addressed both at the stages of design and interpretation of results.[8] More precisely, uses

should be clearly defined *before* designing the test and thus long *before* interpretation and use for decision-making.[9] For example, given their "summative" quality and timing of administration, end-of-year exams are not designed to inform classroom instruction for the assessment year.

Another way of thinking of the uses of assessment would be to categorize them as follows: assessments *for* learning, assessments *as* learning, and assessments *of* learning. Assessments *for* learning enable teachers to use information about students' knowledge and skills to inform teaching and to provide feedback to students to help them monitor and improve their learning. Assessments *as* learning occurs when participating in an assessment not only tracks learning but affects it. Assessments *of* learning monitor knowledge and understanding, as demonstrated by performance on the tests, often in terms of progress toward defined learning goals.

Additionally, assessments should not only measure outcomes (i.e., what students have learned) but also processes (i.e., how teaching and learning is occurring) and "opportunity to learn" constructs. The COVID-19 pandemic in many ways brings to the forefront the importance of understanding and documenting the processes and contexts of learning and the need to account for them in the design and interpretation of assessments.

Table 1 is an abridged representation of the variability and complexity surrounding assessment use as discussed above. It identifies three major contexts (e.g., classrooms and schools, school districts, and state departments of education) where assessments are used,[10] their relative frequency and purpose, and the general intended uses and users for each context. It also identifies the intended uses of the information derived from those assessments and the primary users of that information.

Table 1: Use Context of Assessments

Use Context	Frequency and Purpose	Intended Uses	Intended Users
Classrooms and Schools	• ***Formative***: ongoing during the course of instruction • ***Periodic Summative***: end-of-unit and/or end-of-semester and/ or end-of-year	• Inform instruction • Provide feedback to students • As input to grading	• Teachers • Students • Parents and caregivers • Principals

Table 1 (continued)

Use Context	Frequency and Purpose	Intended Uses	Intended Users
School Districts	• ***Periodic Summative***: monthly; quarterly; semi-annually, as desired	• Feedback and guidance to principals and teachers for improved instruction • Feedback and guidance for school and district leaders on the effectiveness of certain programs, instructional approaches, and curricula • Monitoring of school and district progress • Inform choices for resource allocation	• Teachers • Students • Parents and caregivers • Principals • School district leaders • General public
State Departments of Education	• ***Annual Summative***: comprehensive grade-level coverage on appropriate testing schedule	• Monitoring of system-wide trends through both cohort-referenced and longitudinal growth calculations • Accountability for academic performance • Inform choices for resource allocation	• Teachers • Students • Parents and caregivers • Principals • School district leaders • General public • Policy makers at federal, state, and local levels

INEQUITIES IN PUBLIC EDUCATION

Social and economic inequities affect educational opportunities and outcomes observed in the results of assessment.[11] But efforts to design, administer, and interpret assessments that document disparities in educational achievement need to be sensitive to the ways in which the assessments and their uses may, themselves, perpetuate or exacerbate existing inequities. An underlying predicament is whether the legitimate attempt to *measure* the effects of inequality on education might *cause* further inequality. For example, disadvantaged children's performance on a digital or virtual assessment may be distorted because of inadequate technology or connectivity, which may bias the result and lead to spurious inferences about their learning. Interpretation of results must be conditioned on the possibility that there is no way to fully ensure against differential impacts of the assessment.

2021 END-OF-YEAR STATE SUMMATIVE ASSESSMENTS

Prior to 2002, many states had varied mandatory end-of-year, large-scale assessments. But with the passage of the No Child Left Behind Act of 2001 (NCLB), the federal government mandated measurement of student achievement using annual assessments in

grades 3–8 and once in grades 10–12, reporting of subgroup data at the school level, and school-level indicators based on these annual assessments. States were able to select their assessments and set targets for proficiency, but for accountability purposes had to include yearly increases in proficiency rates for whole schools and identified subgroups. Thus, the use of testing to hold schools, and in some instances teachers, accountable to scores on standardized tests was formalized. These accountability requirements were retained, albeit with some modifications, under ESSA.

In mid-March 2020, because of the pandemic, the U.S. Department of Education (ED) granted temporary waivers to all 50 states, the District of Columbia, the Commonwealth of Puerto Rico, and the Bureau of Indian Education of the U.S. Department of the Interior, which were meant to relieve them from the mandate to administer standardized tests and the associated reporting requirements at the end of the 2019–2020 school year. However, on September 3, 2020, the ED stated that it will not grant waivers of the summative testing requirements for the 2020–2021 school year, citing research[12] that school closures affected the most vulnerable students disproportionately and widened disparities. The ED's policy is based on the argument that assessment data are needed to document learning and educational disparities and to guide decision-making.[13]

Assuming that school districts administer end-of-year summative assessments,[14] below are some caveats concerning the administration, interpretation, and uses of the results.[15] This section concludes with some additional considerations for capturing important student data in the 2020–2021 school year.

Administration

Test administration procedures are developed for an assessment program in large part to reduce measurement error and increase the validity and reliability of the inferences drawn from the assessment. These procedures address numerous factors, such as the timing of test administration, test format (e.g., paper and pencil or digital, multiple choice or other item forms), location and conditions of testing (e.g., remote, in school, in school wearing masks), and implementation of accommodations for test-takers, such as students with disabilities or English learners. The ability for testing sites to adhere to test administration procedures must be examined and contextualized prior to interpreting or using the resulting data. When looking at the 2021 end-of-year, large-scale summative assessments, key test administration procedures to consider are:

- Statutory obligations and constraints. State and school district administration conditions are to some extent mandated by federal and state laws. For instance, as noted above, the federal government mandates the administration of testing and accountability systems. State laws also govern the administration of educational assessments. For example, some states, via statute, policy, or State Board of Education Rule, mandate when testing must occur in the state (i.e., the testing windows), the modality of testing (i.e., paper and pencil or computer-based), the accommodations permitted or mandated (as well as how they are administered), and constraints on remotely proctored examinations (which some states' student privacy statutes would prevent).

- Conditions and contexts of administration. This year, states and districts will likely vary the contexts of administration contemplated for state- or district-level assessments. Although most states are preparing for in-school testing, given that

many districts and schools distributed their electronic devices to students, students taking exams in school buildings may be taking them by different delivery methods (e.g., paper and pencil versus computer). If remote testing does occur, in addition to the tests being in different environments and potentially through different delivery methods, students taking remote tests could encounter connectivity (bandwidth) issues, device malfunctions, and working conditions inconducive to testing (i.e., shared space, distractions by siblings). Moreover, students who have been in remote learning situations could enter schools or examination rooms for the first time on testing days and experience unfamiliar conditions (masks, plexiglass dividers, etc.) that may distort the meaning of their test performance. Social distancing may dictate in-person testing of only a few students at a time, greatly increasing the total time and staffing required to test all students and complicating testing logistics.

Interpretation

The interpretation of the assessment results necessarily requires some type of comparison of scores or other summaries of data. For individual, subgroup, or even school - or district-level interpretations, assessments need to be referenced or compared to prior years, or past performance, or to an absolute standard such as a cut point.[16] Under pre-COVID-19 conditions, comparability concerns were already prevalent and critical to examine. In fact, in early 2020, the NAEd produced a volume, *The Comparability of Large-Scale Educational Assessments: Issues and Recommendations*, addressing how to ensure (or improve) comparability to better interpret test results. Given the disruptions to society and the educational system since March 2020, making valid interpretations[17] from 2020–2021 summative assessment data will be even more difficult. It will be crucial to provide as much *contextual information* as possible when interpreting such data. In addition to the "normal" comparability concerns outlined in the 2020 NAEd volume, there are other high-level considerations that should be addressed in reporting the results of the 2020–2021 assessments:

- Content of instruction. Validity and reliability of inferences from test scores hinge, in most cases, on the extent to which the test is designed to align with standards, curricula, and instruction. Given the pandemic, states, districts, schools, and teachers were forced to prioritize the educational content taught to students. Both last and this school year, some standards and curricula were modified, forsaken, or delayed to a later date or grade level. However, year-end summative assessments in most cases were likely not similarly modified. It is critical for schools and districts to determine what content and skills were actually taught and to provide this contextualization to the test scores.

- Modes of instruction. How content was delivered to students varied not only by states, districts, schools, and classrooms, but also varied *within* these contexts. It also varied over time—some students might have started with remote learning, attended school for a short period of time, and then returned to the remote modality. Moreover, what "remote learning" means significantly varied—for some it was delivered through paper learning packets, others computer-based videos, and for others a mix of synchronous and asynchronous learning. Experts in digital and online pedagogy are quick to emphasize the differences between "emergency remote instruction" and high-quality virtual teaching and learning. Within these diverse environments, instruction varied widely, and the "what" and "how" of this

instruction needs to be reported with finer granularity. For example, synchronous online learning likely varied between and within schools. Similarly, in-school learning varied, and in-school was not "normal" as many students were masked, surrounded by plexiglass, and/or could not use or share manipulatives. For some instruction, students were present in classrooms with some peers but teachers appeared remotely. How content was provided to students will likely affect scores and that context needs to be collected and included with score interpretations.

- Length of instruction. The time devoted to instruction in various subjects also varied widely and needs to be incorporated with the interpretation of the assessment results. Districts and schools adapted to their changing environments and instructional modes by adjusting the number of days of schooling (including closing schools due to health risks) and by adjusting the length of the instructional day, which affected both the "amount" and effectiveness of learning. Assessments would need to capture these external sources of variation.

- Conditions and contexts of administration. While most states are planning for in-person administration of assessments, some may permit remote examinations; as described above, even these examinations will likely take different forms. The conditions and contexts of administration are likely nonrandom and could affect claims concerning comparability and other important components of assessment interpretation.

- Participation rates. Not all students will take end-of-year assessments. While we have yet to see a national opt-out movement, some research already reports that a majority of parents support cancelling the 2020–2021 end-of-year summative assessments.[18] It is likely that some parents, caregivers, and students will choose not to have their children (or themselves) return to campus—if they are in remote learning—simply to enable testing. Additionally, if those opting out are nonrandomly distributed and include a larger percentage of historically marginalized or disadvantaged students and others who are relatively less engaged in schooling because of the pandemic, the interpretation of assessment results will become more challenging, with test results not supporting valid or reliable inferences about performance. On a final note, ESSA requires that 95% of all students and 95% of all student subgroups participate in the end-of-year state assessments. With likely lower participation rates, this factor is another comparability and interpretive dimension that must be contextualized (i.e., who did and did not get tested, and why?).

- Social and emotional well-being. While assessment conditions for some students can be stressful and anxiety-producing in "normal" times, the pandemic is likely to make things even worse. Some students will have concerns about health risks and physical safety in school buildings. Moreover, unusual testing conditions, such as mask wearing, distancing, and barricades to prevent the spread of the virus can increase stress, which further compromises the validity and reliability of assessment data. Students' social and emotional well-being prior to and during testing will likely be reflected in their scores but are difficult to account for accurately. Again, interpretations of assessment results must be sensitive to contextual determinants of student well-being.

- Opportunity to learn. The opportunity to learn (OTL), with respect to assessments, has been conceptualized as the opportunity to learn what is tested.[19] It includes, among other factors, school resources, access to the curriculum, time allocated for instruction, quality of instruction, coverage of the curriculum, access to culturally responsive teaching and curriculum, disciplinary and exclusionary practices, teaching to the test, appropriate identification and services for students with disabilities, familiarity with item formats and tools used for assessments, and students' preparedness to participate in learning.[20] Dissimilar OTL presents a threat to comparability and interpretation. However, if an assessment is used purely to describe students' current levels of achievement, without implying attribution to the various factors that explain those results, then OTL considerations may be less necessary.

Use Cases

Above we address the caveats needed to make any interpretations from the 2020–2021 summative assessments. Here, we address specific "use" cases of the 2020–2021 end-of-year summative assessments. Assessments are designed and validated for specific uses. ESSA-mandated tests are meant to be designed to measure student achievement and these data are then used for ESSA-mandated accountability.[21] ESSA also requires, for accountability purposes, that data be disaggregated to the subgroup level. In addition to federal requirements,[22] states also have various other mandates surrounding the uses of educational assessments, including for grade promotion, teacher evaluations, high school graduation, certain student grading, and ranking or rating schools.

- Accountability. ESSA requires school-level accountability for student achievement and mandates that academic achievement as well as another academic indicator, which almost every state measures as academic progress (both based on assessments) be indicators of student achievement. Putting aside the measurement complications of academic progress since there are no 2019–2020 assessment data to measure student growth, it is difficult, if not impossible, to untangle the effects of the pandemic from school performance. Holding schools accountable for outcome data during this pandemic seems unfair, to say the least, and for some educators borders on the unconscionable. There is a compelling argument that the 2020–2021 summative assessments cannot accurately be used to rate or categorize schools (or teachers) and even reporting such data may invite unsupportable interpretations by policy makers and members of the public.[23] In short, using assessments for accountability, which is always fraught with complexity and controversy, becomes even more problematic during (and after) the pandemic. (For more on accountability, see "Beyond 2020–2021 Assessments" below.) Similarly, states may need to rethink using assessments for certain state-mandated accountability decisions (e.g., promotion, graduation, grading, teacher evaluation, ranking and rating schools).

- Educational inequities. NCLB was the first time states were federally required to report and account for subgroup assessment results. As a result of this NCLB mandate, inequities among subgroups at the school building level were identified and highlighted for the first time. ESSA, NCLB's successor, requires, as did NCLB, that school-level accountability data be reported by the following subgroups: economically disadvantaged students; racial/ethnic groups; students with disabilities as defined by the Individuals with Disabilities Education Act (IDEA);

and English learners. This requirement has highlighted not only the large inequities across subgroups across states and school districts, but also within schools. There is little doubt that if *all* students are tested, the 2020–2021 end-of-year summative assessments will show that the COVID-19 pandemic has exacerbated already existing inequities.[24] If testing of *all* students occurred using high-quality, end-of-year assessments, they would provide generalized, aggregated information to help gauge the impact of the pandemic on learning and growth and could be used to inform 2021–2022 school planning and resource allocation. However, test participation will be nonrandom, with students most harmed by the pandemic and thus with the most academic and social and emotional loss likely being the ones not tested. Finally, spending limited instructional or in-person time by conducting large-scale assessments may not be the best way to measure this exacerbation of educational inequities. (For more information, see "Additional Considerations for Gathering 2020–2021 Data" below.)

Additional Potential Consequences of Administering 2020–2021 End-of-Year Summative Assessments

The potential negative consequences of the administration of the 2020–2021 assessments should be mitigated insofar as possible; caveats and contexts should be part of plans to report scores. Public reporting of assessment data often feels punitive to school personnel, and potential unintended consequences and misuse that come with public reporting should be anticipated and addressed. Moreover, opportunity costs associated with trying to test students who have recently returned to school or are not in school must be minimized. As research has demonstrated, those students most in need of high-quality instruction spend the most time in test preparation, and losing limited instructional time is more consequential now than ever before. Results should not be viewed with a "deficit" mindset (i.e., the mindset of who can and cannot learn based on the dominant cultural stereotypes), but instead as an opportunity to provide supports and services.[25]

Additional Considerations for Gathering 2020–2021 Data

Putting aside the 2020–2021 end-of-year summative assessments, there are assessments and data that will be important to collect and use during the 2020–2021 school year to enhance teaching and learning and allocate resources equitably. Here are some examples:

- Opportunity to learn. As noted above, for valid and reliable interpretations and uses of assessments, the assessments must be appropriately contextualized. For instance, it must be understood how much of the curriculum was covered, how the material was imparted (in-person, remotely, synchronous, asynchronous), and the composition of students' learning environments. The instructional delivery mode needs to be further parsed as, for example, synchronous online learning likely widely varies and these differences should be, to the extent possible, examined. It is important to understand if students were engaging in remote learning and at the same time caring for younger siblings, suffering from food insecurity, struggling socially and emotionally, sharing limited technology devices, struggling to access Wi-Fi, or learning in abusive environments. If the 2020–2021 exams were to be administered, this contextual information would be necessary for interpretation.[26] Moreover, regardless of contextualizing the summative examinations, gathering OTL data is critical to highlight educational inequities and to allocate resources and supports accordingly.[27]

- Classroom- and district-level assessments. Classroom- and district-level assessments provide information that can support the improvement of teaching and student learning. Now, they are more important than ever to help district and school leaders and classroom teachers monitor the state of student learning, including major inequities resulting from the pandemic, and target resources to address those students most in need of support. (For more on the formative and summative use of such assessments, see the next section.)

BEYOND 2020–2021 ASSESSMENTS

In addition to addressing the COVID-19 crisis, education leaders, teachers, students, and caregivers have also been facing what many describe as three additional pandemics—America's reawakening to realities of racial injustice and violent extremism, an economic recession that shows no signs of significant recovery in the near term, and a climate crisis. These crises are both pushing us to deal with the immediacy of needs and also challenging us to rethink what a "normal" return means, especially for vulnerable students (e.g., English learners; members of historically marginalized, disadvantaged, and underserved groups; and students with disabilities). There is considerable momentum to seize the current moment and correct fundamental flaws in the education system, among them the tolerance of low proficiency levels in literacy and math, high rates of suspension and expulsion, over-identification (and misclassification) of students for special education, and low rates of high school completion. Educators, policy makers, scholars, parents and caregivers, and students are all grappling with the question of what is important to know? And consequently, what is important to measure?

As we continue to define what to measure, we must ensure that assessments reliably and validly measure the "what." Below are goals for assessments that emerged during the roundtable conversation to consider as we think about assessments and their uses to help improve teaching and learning by informing decisions about students; teachers, curricula, programs, and schools; funding; and policy.

Develop and Implement Culturally and Racially Responsive, Curriculum-Embedded, Balanced Assessment Systems. Assessments, including classroom-, district-, and state-level assessments, need to be integrated with standards, curricula, and instruction. Assessments are a critical part of the educational system, and they need to both be themselves as free as possible from racial and cultural biases and similarly be aligned to such standards, curricula, and instruction.

- An assessment system is ***balanced*** when: the various types of assessments in the system are coherently linked through a clear specification of learning targets, they comprehensively provide multiple sources of evidence to support educational decision-making, and they continuously document student progress over time.[28] A balanced assessment system demands that ***states and districts work together*** to coordinate the assessments used at classroom, district, and state levels to reflect these principles of coherence, comprehensiveness, and continuity.
- Balanced assessment systems include aligned formative and summative assessment activities. ***Classroom-level formative assessment activities*** can include short, quick activities embedded in curricula to provide teachers and students up-to-the-minute information about the outcomes of teaching and learning. These activities can rapidly indicate actionable areas of strength and weakness as well as skill and knowledge development. The formative use of

assessment for and as learning is critical to a balanced system.[29] Throughout the year, ***classroom- and district-level summative assessments*** provide small to intermediate scale tests aligned to the units of instruction. Such curricula-aligned assessments (often accompanied by student projects, portfolios, capstones, and performance tasks) provide actionable information about student attainment and progress for teachers, students, and parents and caregivers. At the school and district level, such assessments can also assist in identifying teacher professional development needs as well as informing equitable resource allocation. When a balanced assessment system is working, the ***state-level summative assessments*** will play important roles in overall district- and school-level monitoring and resource allocation. Such large-scale assessments of learning can help identify differences among groups and can inform accountability measures (for more information, see "Reframe Test-Based Accountability from a Deficit Lens to an Improvement Perspective" below), but typically cannot provide the more time-sensitive information needed to improve ongoing teaching and learning. Consequently, using ***formative assessments*** to guide instructional practices and a balanced portfolio of classroom-, district- and state-level ***summative assessments*** to monitor achievement and guide resource allocation will likely increase the contribution of a system of assessments to the promotion of more equitable education.

Implement Equitable Educational Assessments. As noted above, inequities in public education are multifaceted (e.g., including socioeconomic factors, instructional quality, linguistic differences) and impact students' performance on assessments, and ultimately, decisions based on those assessments. Moreover, assessments themselves are potentially subject to inequities in design, content, OTL, and language choices. Here, we stress the need for not only equity in education but also for ***equitable educational assessments***. An equitable educational assessment system is:

- ***Fair***: Fair assessments are sensitive to the characteristics of different groups being assessed and thereby, where appropriate, reflect diversity in the design and delivery of the assessment and the reporting of assessment results.
- ***Accurate***: Accurate measurement occurs when measurement error is minimized as equally as possible for all groups of test-takers.
- ***Valid***: Equitable measures are aligned and validated with their interpretations and uses.[30]

Communicate Clearly (and Often) the Intended Purposes and Uses of Particular Assessments as Well as Any Relevant Context. People want tests to provide simple answers to complex questions. Instead, we must continue to emphasize the intended interpretations and uses of each particular assessment. We also must contextualize assessment results, emphasize measurement error and uncertainty, and warn against unwarranted causal attributions. For example, state leaders should meet with staff from media outlets well before test results are produced to provide them with a framework for interpreting the results.

Reframe Test-Based Accountability from a Deficit Lens to an Improvement Perspective. Accountability now is viewed as a punishment or a sanction. Federal accountability requires the labeling of schools and negative consequences. Instead, assessment and accompanying accountability should lead to improving schools and districts' ability to provide equitable OTLs and to help students to take maximal advantage

of those opportunities. Evidence from research and practice cautions against the simplistic notion that the threat of sanctions based on tests creates incentives for genuine improvement of teaching.[31] Alternatively, the design and use of assessments could be based on the proposition that teachers want information to guide their work. As such, assuming that most teachers want to do more for their students, assessment data, paired with best practices, could fuel improvement. Accountability should be the guide to improvement, and improvement-based accountability provisions should be piloted.

Measure Opportunity to Learn. Assessments alone cannot be expected to identify discrepancies in OTLs. Large-scale assessments do highlight educational gaps and inequities suggesting OTLs. However, fine-grain measures of OTL are necessary to understand why and how the gaps exist as well as to provide a roadmap to addressing educational inequities.[32]

Expand Assessment Literacy. Assessments are only useful if those who could benefit from the information can access, interpret, and use the information to improve teaching and learning. Recognizing that appropriately educating all who interpret and make use of educational testing data is no small task, we offer a few suggestions. First, we need to ensure that the right people quickly gain access to and use testing data. Second, we need to ensure that teachers, administrators, parents and caregivers, and students are educated in how to interpret and use assessments to further teaching and learning and create equitable educational opportunities. For teachers, this may result in professional development and in-service opportunities. Like all aspects of education, parents and caregivers need to be seen as integral partners in using assessments to further learning. Finally, it is critical that policy makers and media outlets are provided with a framework and context to understand, interpret, and report results.

Examine the Equity Concerns Inherent in Other Assessments. While the NAEd roundtable focused on formative and summative assessments, there are additional assessments, particularly those used for diagnostic and classification purposes that are rife with equity concerns. Assessments are used to diagnose (e.g., disabilities), classify (e.g., English learners), place and assign (e.g., gifted and talented, advanced placement), promote and demote, and certify and graduate students. Prior to the COVID-19 pandemic, some such assessments raised validity and equity concerns. We imagine that because of the pandemic, these assessments may compound inequities and students who have historically benefited from such assessments are more likely to benefit now and those historically harmed will likely be more harmed. Thus, these assessments need to be closely examined and refined to ensure that they are valid measures of their intended purposes and do not instead further exacerbate educational inequities.

Encourage Innovation and Flexibility. Ideally, systems of assessment would serve the improvement of institutions, the improvement of teaching and learning, the improvement of teachers, and the improvement of students. To accomplish this, we need to encourage appropriate, mindful, and documented flexibility and innovation to see what works and when, and then the results can be used to encourage some degree of uniformity at a more macro level. At the federal level, through the reauthorization of ESSA, the federal government should encourage innovation in assessment and accountability, including approaches that look beyond testing per se. ESSA should provide waivers and fund pilots for states to appropriately experiment with assessment techniques and provide feedback

on successful (and unsuccessful) elements of innovative assessment systems and accountability measures. This innovation should not be in lieu of working to enhance equitable and fair educational opportunities but as a mechanism to further equitable opportunities.

Address Ongoing COVID-19 Implications. COVID-19 implications will be felt for years, and we must continue to attempt to measure these implications on both academic and social and emotional learning and provide supports to address them. Moreover, we will have a generation of children impacted by the COVID-19 pandemic who will lack benchmark assessments, have inconsistent measures, or for a variety of factors stated above have summative assessment measures impacted by OTL or other contextual variables. We must be vigilant to monitor and address the COVID-19 legacy, particularly for our historically disadvantaged children.

CONCLUSION

Assessments, if used properly, can help us to mitigate the impacts of the COVID-19 pandemic for years to come. If used improperly, assessments may waste precious instructional time and resources, worsen inequities, reinforce misperceptions as to sources of inequity, and impede sound education policy. While most people agree that critical data are needed to measure academic knowledge, the "what" and "how" continue to afflict us. Thus, we encourage further discussions among educators, researchers, policy makers, and the general public to work toward making sure educational assessments are part of a system to further teaching and learning and to further the pursuit of equity.

List of Useful Resources

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

Alexander, L., James, H. T., & Glaser, R. (1987). *The nation's report card: Improving the assessment of student achievement*. Washington, DC: National Academy of Education.

Baghian, J. (2021). Assessment data can help us build back better. *Education Next Forum*.

Bennett, R. (2020). *Interpreting test-score comparisons*. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 227–235). Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1.

Berman, A. I., Feuer, M. J., & Pellegrino, J. W. (2019). What use is educational assessment? *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 8–20. https://doi.org/10.1177/0002716219843871.

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.). (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90. https://doi.org/10.1177/003172171009200119.

Boyer, M., Dadey, N., & Keng L. (2020, September). *Statewide summative assessment in spring 2021: A workbook to support planning and decision-making*. Dover, NH: National Center for the Improvement of Educational Assessment.

Connecticut State Department of Education. (2020, June 29). *Sensible assessment practices for 2020-21 and beyond*.

Council of Chief State School Officers. (2020). *Restart & recovery: Assessments in spring 2021*.

DePascale, C., & Gong, B. (2020). *Comparability of individual students' scores on the "same test"*. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 25–48). Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1.

DeVos, B. (letter, September 3, 2020). Key policy letters signed by the Education Secretary or Deputy Secretary.

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020, June). *COVID-19 and student learning in the United States: The hurt could last a lifetime*. McKinsey & Company.

List of Useful Resources

Glaser, R., Linn, R., & Bohrnstedt, G. (1997). *Assessment in transition: Monitoring the nation's educational progress*. Washington, DC: National Academy of Education.

Gordon, E. W. (1995). Toward an equitable system of assessment. *The Journal of Negro Education*, *64*(3), 360–372.

Haertel, E., & Ho, A. (2016). Fairness using derived scores. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement (1st ed.)*. New York: Routledge. https://doi.org/10.4324/9781315774527.

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. https://doi.org/10.1080/0969594X.2015.1060192.

Keng, L., Boyer, M., & Marion, S. F. (2020). Into the unknown: Assessment considerations for spring 2021. *Educational Measurement: Issues and Practice, 39*(3), 53–59. http://dx.doi.org/10.1111/emip.12362.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020, December 3). *How is COVID-19 affecting student learning? Initial findings from fall 2020*. Brookings, Brown Center Chalkboard.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher, 49*(8), 549–565. https://doi.org/10.3102/0013189X20965918.

Marion, S. (2020, October). *Using opportunity-to-learn data to support educational equity*. Dover, NH: National Center for the Improvement of Educational Assessment.

Marion, S., Gong, B., Lorié, W., & Kockler, R. (2020, July). Restart & recovery: Assessment consideration for fall 2020. Council of Chief State School Officers.

Marion, S. F., Gonzales, D., Wiener, R., & Peltzman, A. (2020). *This is not a test, this is an emergency: Special considerations for assessing and advancing equity in school-year 2020–21*. National Center for the Improvement of Educational Assessment (www.nciea.org) and The Aspen Institute (www.aspeninstitute.org/education).

Marion, S. & Shepard, L. (2021). Focus on instructional and intervention, not testing, in 2021. *Education Next Forum*.

Mislevy, R. J. (2019). Advances in measurement and cognition. In A. I. Berman, M. J. Feuer, & J. W. Pellegrino (Eds.) *What use is educational assessment? The ANNALS of the American Academy of Political and Social Science*, *683*(1), 164–182. https://doi.org/10.1177/0002716219843816.

Moss, P. A., Pullin, D. C., Gee, J., Haertel, E. H., & Young, L. J. (2008). *Assessment, equity, and opportunity to learn*. Cambridge, UK: Cambridge University Press.

List of Useful Resources

National Academy of Education. (2009). *Education policy white paper on standards, assessments, and accountability*. L. Shepard, J. Hannaway, & E. Baker (Eds.). Washington, DC: Author.

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25389.

National Education Association. (2003). *Balanced assessment: The key to accountability and improved student learning*. Washington, DC: Author.

National Research Council. (2011). *Incentives and test-based accountability in education*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12521.

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

National Research Council & National Academy of Education. (2010). *Getting value out of value-added: Report of a workshop*. Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability. H. Braun, N. Chudowsky, & J. Koenig (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

New Mexico Public Education Department. (2020).

- *Reentry guidance*.
- *Using multiple measures & formative practice to identify learning needs: reentry guidance*.
- *Instructional acceleration*.
- *New Mexico's digital distance learning recommendations: 2020 & beyond*.

Shepard, L. A. (2020, December 16). *Testing students this spring would be a mistake*. Education Week.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14.

Silver, D., & Polikoff, M. (2020, November 16). *Getting testy about testing—K–12 parents support canceling standardized testing this spring. That might not be a good idea*. The 74.

Singer, J. D., Braun, H. I., & Chudowsky, N. (Eds.). (2018). *International education assessments: Cautions, conundrums, and common sense*. Washington, DC: National Academy of Education.

Soland, J., Kuhfeld, M., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020, May 27). *The impact of COVID-19 on student achievement and what it may mean for education.* Brookings, Brown Center Chalkboard.

U.S. Congress, Office of Technology Assessment (1992). *Testing in American schools: Asking the right questions*. Washington, DC: U.S. Government Printing Office.

Roundtable Panelists

Educational Assessments in the COVID-19 Era and Beyond

Thursday, December 10, 2020

Randy E. Bennett
Norman O. Frederiksen Chair in Assessment Innovation
Research & Development Division
Educational Testing Service

Debbie Durrence
Chief Data Officer
Gwinnett County Public Schools

Michael Feuer (Moderator) *
Dean and Professor, Graduate School of Education and Human Development
The George Washington University
Immediate Past President
National Academy of Education

Ajit Gopalakrishnan
Chief Performance Officer
Connecticut State Department of Education

Edward Haertel *
Jacks Family Professor of Education, Emeritus
Stanford University

Gerunda Hughes
Professor Emeritus
Howard University

Peter Leonard
Director of Student Assessment
Chicago Public Schools

Scott Marion
President and Executive Director
National Center for the Improvement of Educational Assessment

Kent McGuire
Program Director of Education
William and Flora Hewlett Foundation

Scott Norton
Deputy Executive Director of Programs
Council of Chief State School Officers

James W. Pellegrino *
Liberal Arts and Sciences Distinguished Professor & Co-director of the Learning Sciences Research Institute
University of Illinois at Chicago

Kenneth A. Shores
Assistant Professor, School of Education
University of Delaware
UD Center for Research in Education and Social Policy

Jim Soland
Assistant Professor of Research, Statistics, and Evaluation
University of Virginia
Associated Research Fellow
NWEA

Lynn Vasquez
Division Director of Assessment and Learning Management Systems
New Mexico Public Education Department

Vince Verges
Assistant Deputy Commissioner of the Division of Accountability, Research, and Measurement
Florida Department of Education

** denotes roundtable planning committee member*

Steering Committee Members

Michael Feuer (Chair)
Dean and Professor, Graduate School of Education and Human Development
The George Washington University

Hyman Bass
Samuel Eilenberg Distinguished University Professor of Mathematics & Mathematics Education
University of Michigan

Dorothy Espelage
William C. Friday Distinguished Professor of Education
University of North Carolina at Chapel Hill

Gloria Ladson-Billings
Professor Emerita, Department of Curriculum & Instruction
University of Wisconsin-Madison

Susanna Loeb
Director, Annenberg Institute at Brown University
Professor of Education and International and Public Affairs
Brown University

Annemarie Sullivan Palincsar
Chair of Educational Studies
Jean and Charles R. Walgreen Jr. Professor of Reading and Literacy
Arthur F. Thurnau Professor
University of Michigan

William F. Tate IV
Provost and Executive Vice President of Academic Affairs
Distinguished Professor of Sociology and Family and Preventive Medicine
University of South Carolina

Frank Worrell
Professor and Director, School Psychology
Faculty Director, Academic Talent Development Program
Faculty Director, California College Preparatory Academy
Affiliate Professor, Psychology
University of California, Berkeley

Stanton Wortham
Inaugural Charles F. Donovan, S.J., Dean
Boston College

NAEd Staff

Amy I. Berman
Deputy Director

Dian Dong
Senior Program Officer

Endnotes

[1] This excerpt comes from a "Dear Friends and Colleagues" letter that Dr. Teresa Thayer Snyder posted to Facebook on December 6, 2020. It was widely shared including in full on Diane Ravitch's blog on December 12, 2020.

[2] The Editorial Board (2021, January 2). The wreckage Betsy DeVos leaves behind: The Education Department lies in ruins right when it's needed most. *The New York Times*.

[3] See National Research Council & National Academy of Education. (2010). *Getting value out of value-added: Report of a workshop*. Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability. H. Braun, N. Chudowsky, & J. Koenig (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; National Academy of Education. (2009). *Education policy white paper on standards, assessments, and accountability*. L. Shepard, J. Hannaway, & E. Baker (Eds.). Washington, DC: Author; Glaser, R., Linn, R., & Bohrnstedt, G. (1997). *Assessment in transition: Monitoring the nation's educational progress*. Washington, DC: National Academy of Education; Alexander, L., James, H. T., & Glaser, R. (1987). *The nation's report card: Improving the assessment of student achievement*. Washington, DC: National Academy of Education.

[4] See Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.). (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1; Berman, A. I., Feuer, M. J., & Pellegrino, J. W. (2019). What use is educational assessment? *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 8–20. https://doi.org/10.1177/0002716219843871; Singer, J. D., Braun, H. I., & Chudowsky, N. (Eds.). (2018). *International education assessments: Cautions, conundrums, and common sense*. Washington, DC: National Academy of Education.

[5] The words *assessment* and *test* are used throughout this report, and though to some extent they are interchangeable, they do have different meanings. *Assessment* is more general, conveying the idea of a process providing evidence of quality. *Assessment* covers a broad range of procedures to measure teaching and learning. A *test* is one product that measures a particular set of objectives or behavior. See Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). *Introduction: Framing the issues*. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 9–24). *Comparability of large-scale educational assessments: Issues and recommendations*. Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1.

[6] While not all uses of testing are justifiable, it is important to recognize them so that they can be addressed when attempting to expand assessment literacy. (For more information see "Beyond 2020–2021 Assessments" below.)

[7] Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.). (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1; Berman, A. I., Feuer, M. J., & Pellegrino, J. W. (2019). What use is educational assessment? *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 8–20. https://doi.org/10.1177/0002716219843871; Connecticut State Department of Education. (2020, June 29). *Sensible assessment practices for 2020–21 and beyond*; National Research Council. (2001). National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press; U.S. Congress, Office of Technology Assessment. (1992), *Testing in American schools: Asking the right questions*. Washington, DC: U.S. Government Printing Office.

[8] Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.). (2020). *Comparability of large-scale educational assessments: Issues and recommendations*. Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1; AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). (2014). *Standards for educational and psychological measurement*. Washington, DC: AERA.

[9] Haertel, E., & Ho, A. (2016). Fairness using derived scores. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement (1st ed.)*. New York: Routledge. https://doi.org/10.4324/9781315774527.

[10] There are of course additional contexts where assessments are used. For example, the federal government administers the National Assessment of Educational Progress (NAEP), which is given to a representative sample of students across the country to garner national, state, and some urban district measures of what students know across various subject areas.

Endnotes

There also are various international assessments in which the United States participates.

[11] For a discussion of the different meanings of inequity and inequality, see, e.g., National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25389.

[12] Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020, June). *COVID-19 and student learning in the United States: The hurt could last a lifetime*. McKinsey & Company.

[13] DeVos, B. (letter, 2020, September 3). Key policy letters signed by the Education Secretary or Deputy Secretary.

[14] In December 2020, *Education Week* published an opinion piece in which Lorrie Shepard presented many of the concerns with administering and using 2020–2021 end-of-year summative assessments. See, Shepard, L. A. (2020, December 16). Testing students this spring would be a mistake. *Education Week*. Shepard also responded to numerous civil rights advocacy groups urging 2020–2021 summative exams. See Civil Rights Organizations. (letter, 2020, November 20). Letter to Deputy Assistant Secretary Ryder, U.S. Department of Education. See also Baghian, J. (2021). Assessment data can help us build back better. *Education Next Forum*; Marion, S. & Shepard, L. (2021). Focus on instructional and intervention, not testing, in 2021. *Education Next Forum*.

[15] Boyer, M., Dadey, N., & Keng L. (2020, September). *Statewide summative assessment in spring 2021: A workbook to support planning and decision-making*. Dover, NH: National Center for the Improvement of Educational Assessment; Council of Chief State School Officers. (2020). *Restart & recovery: Assessments in spring 2021*; Keng, L., Boyer, M., & Marion, S. F. (2020). Into the unknown: Assessment considerations for spring 2021. *Educational Measurement: Issues and Practice, 39*(3), 53–59. http://dx.doi.org/10.1111/emip.12362.

[16] Bennett, R. (2020). *Interpreting test-score comparisons*. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 227–235). Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1.

[17] "The validation of a score interpretation involves an investigation of whether the scores mean what they are supposed to mean, and the interpretation is said to be valid if claims inherent in the interpretation are supported by appropriate evidence." Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. https://doi.org/10.1080/0969594X.2015.1060192.

[18] Silver, D., & Polikoff, M. (2020, November 16). *Getting testy about testing—K–12 parents support canceling standardized testing this spring*. That might not be a good idea. The 74.

[19] DePascale, C., & Gong, B. (2020). *Comparability of individual students' scores on the "same test"*. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 25–48). Washington, DC: National Academy of Education (citing Moss, P. A., Pullin, D. C., Gee, J., Haertel, E. H., & Young, L. J. (2008). *Assessment, equity, and opportunity to learn*. Cambridge, UK: Cambridge University Press.). https://doi.org/10.31094/2020/1.

[20] Id.

[21] Bennett, R. (2020). *Interpreting test-score comparisons*. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 227–235). Washington, DC: National Academy of Education. https://doi.org/10.31094/2020/1; Berman, A. I., Feuer, M. J., & Pellegrino, J. W. (2019). What use is educational assessment? *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 8–20. https://doi.org/10.1177/0002716219843871.

[22] For many reasons identified in this section, including concerns about representativeness, comparability, reliability, and validity, the National Center for Education Statistics (NCES) and the National Assessment Governing Board (NAGB) postponed the 2021 administration of the National Assessment of Educational Progress (NAEP) in reading and mathematics. "At the Governing Board's Nov. 19-20 meeting, NCES presented compelling data, which convinced Board members that COVID-19 related conditions prevent NCES from administering NAEP safely to a sufficient and representative sample, and reporting results in a valid and reliable manner consistent with NCES' statistical standards and the NAEP Authorization Act. Thus, the Governing Board believes a 2022 administration of NAEP reading and mathematics at grades 4 and 8 would be more likely to provide valuable—and valid—data about student achievement in the wake of COVID-19 to support effective policy, research, and resource allocation" (NAGB. (2020, November 25). *Governing Board statement on postponement of NAEP 2021*.).

Endnotes

[23] Shepard, L. A. (2020, December 16). *Testing students this spring would be a mistake*. Education Week; Marion, S. F., Gonzales, D., Wiener, R., & Peltzman, A. (2020). *This is not a test, this is an emergency: Special considerations for assessing and advancing equity in school-year 2020–21*. National Center for the Improvement of Educational Assessment (www.nciea.org) and The Aspen Institute (www.aspeninstitute.org/education).

[24] Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher, 49*(8), 549–565. https://doi.org/10.3102/0013189X20965918.

[25] Additionally, there are significant logistical and operational challenges to administering end-of-year assessments including staffing needs and concerns, distancing requirements, protective equipment requirements, in-school device availability (e.g., due to providing in-home devices), safe handling of materials (e.g., papers, pencils), security, and remote proctoring requirements and costs.

[26] We recognize that OTL data can be difficult to collect and coherently and uniformly report, and that the COVID-19 pandemic has made such collection more difficult.

[27] Marion, S. (2020, October). *Using opportunity-to-learn data to support educational equity*. Dover, NH: National Center for the Improvement of Educational Assessment.

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

[29] See, e.g., New Mexico Public Education Department (2020) documents "Using Multiple Measures and Formative Practice to Identify Learning Needs," "Reentry Guidance," and "Instructional Acceleration."

[30] Mislevy, R. J. (2019). Advances in measurement and cognition. In A. I. Berman, M. J. Feuer, & J. W. Pellegrino (Eds.) *What use is educational assessment? The ANNALS of the American Academy of Political and Social Science*, *683*(1), 164–182. https://doi.org/10.1177/0002716219843816.

[31] National Research Council. (2011). *Incentives and test-based accountability in education*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12521.

[32] Marion, S. F., Gonzales, D., Wiener, R., & Peltzman, A. (2020). *This is not a test, this is an emergency: Special considerations for assessing and advancing equity in school-year 2020–21*. National Center for the Improvement of Educational Assessment (www.nciea.org) and The Aspen Institute (www.aspeninstitute.org/education); National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25389.

Additional Information

Suggested Citation:

National Academy of Education. (2021). *Educational assessments in the COVID-19 era and beyond*. Washington, DC: Author.

For inquiries, contact Amy Berman, Deputy Director (aberman@naeducation.org), or Dian Dong, Senior Program Officer (ddong@naeducation.org).

This project was supported by a grant from the Spencer Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the National Academy of Education and do not necessarily reflect the views of the Spencer Foundation.