RESTART & RECOVERY:
ALTERNATIVE APPROACHES TO STATE TESTING IN 2021

CCSSO

Council of Chief State School Officers ™

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, Bureau of Indian Education, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

EXECUTIVE SUMMARY

The global pandemic has disrupted many things across education systems in this country, including the traditional assessment tools that states have used to measure student achievement. As the Council of Chief State School Officers (CCSSO) noted in a recent public statement, despite these challenges, state education leaders remain deeply committed to knowing where students are academically and using data to inform decision-making. However, the reality of how states can assess students this school year and report meaningful data may need to look different, depending on each state's context and the impact that COVID-19 has had on its schools. According to CCSSO, "Today, states are moving forward and exploring how to administer their statewide summative assessment to as many students as possible this spring or exploring other similar important measures."

Given the unresolved, ongoing conditions caused by the COVID-19 pandemic in each state and local community, only one thing is clear: there are no easy answers. It is increasingly apparent that states may need to consider solutions that include a range of unconventional alternatives to measure the academic progress of as many students as possible, depending on the conditions on the ground in their state. CCSSO commissioned this paper to inform the thinking of state education leaders across the country; CCSSO does not endorse one option over another but is working to ensure states are aware of the many alternatives that could be possible or should be considered. The authors have taken care to identify the tradeoffs various alternatives entail, so that policymakers and practitioners, working together, can make more informed choices.

Prior to exploring potential solutions, this paper will outline the three major categories of challenges to statewide assessments and valid score interpretations in 2021.

EQUITY

The COVID-19 disruptions will exacerbate existing inequities in terms of both learning and assessment. It is a moral imperative to evaluate any proposed assessment solution in relation to the degree to which a given solution will meet the needs of all students.

UTILITY

Given the COVID-19 disruptions to in-person learning beginning in spring 2020, it is reasonable to question whether the results from large-scale state tests can be used meaningfully. Any proposed solutions will need to address whether 2021 assessment scores will serve the intended purposes and uses in meaningful ways.

CAPACITY

Education leaders may need to face a situation in which they cannot safely test students inside school buildings, which raises concerns about how to maintain standardized test-administration conditions.

Indeed, in light of these three potential challenges, some education leaders may be reluctant to use valuable financial and personnel resources to support state summative testing.

With these considerations in mind, this paper offers the following approaches to testing in 2020-2021. Note that some of these ideas may not be consistent with the statutory requirements of the Every Student Succeeds Act (ESSA). However, Secretary DeVos' September 2020 letter to chief state school officers invited state leaders to consider alternatives potentially outside the lines of current federal policy.

EXECUTIVE SUMMARY (CONT.)

There are four general frameworks in which to consider the proposed alternatives:

ACCESS & SUPPORT

How can states bolster the capacity of schools and districts to address the needs of all students, particularly those who may be most vulnerable? While this is not a testing alternative, it is a critical overarching concern.

DESIGN

How might states select or revise the assessment instrument?

ADMINISTRATION

How can states modify the manner in which the assessment is implemented?

INTERPRETATION & USE

How should states help stakeholders understand and use the results in an appropriate and defensible manner?

Education leaders are strongly urged to **collect and report opportunity-to-learn (OTL) information** by conducting a comprehensive needs assessment that is similar to the Center for Assessment's suggestions for collecting and reporting opportunity-to-learn information.[1,2] Further, if states administer assessments in 2021, it will be important to contextualize the reported results by linking them to key OTL information.

Decouple assessment from accountability, because using summative assessments in spring 2021 as part of a state's status quo accountability system is likely to be problematic. Note that some sort of permission from the U.S. Department of Education will be required for this option. Nevertheless, the impediments to providing adequate opportunities to learn, as well as the difficulty in producing scores that can be meaningfully compared with prior years' performance data, would seem to warrant eliminating high-stakes, test-based accountability for students, educators, and schools in 2020-2021. CCSSO voiced support for this action in a statement made on December 17, 2020.

The authors considered several approaches to testing in 2020-2021 which would increase efficiency by **reducing the length and/or number of assessments**. States may wish to consider using one or more of these approaches to reduce the burden of testing. Strategies such as **grade-band testing** offer the benefit of (a) allowing for more socially-distant testing within a reasonable testing window and (b) producing the most cost-savings. The obvious drawback of sampling students in this fashion is the considerable loss of information on student achievement. Not only does it preclude trend comparisons for grades no longer tested, it complicates — and possibly prohibits — calculations of student growth.

Expanding testing windows is perhaps the least disruptive alternative to administration. Most states currently have a fairly large window within which to determine the timing and duration of testing. Nevertheless, expanding the testing window permits more flexibility for "cycling" students through the test administration, which mitigates safety concerns by testing smaller, more spread-out groups of students. If combined with test-length reductions, this may further bolster capacity for small-group testing.

There are, however, at least two concerns associated with this alternative. First, any time the test window is expanded, there is an increased risk that test content will be shared. Second, if the testing window is expanded too much, there may be a pronounced difference in opportunity to learn for early testers, as compared with later testers. In short, the window should not be so broad that students testing early are denied adequate instruction.

Finally, state leaders should work closely with their assessment experts, technical advisors, and key stakeholders (e.g., district leaders, critical policymakers) to **begin planning as soon as possible for multiple contingencies in this school year**. Perhaps most importantly, state leaders need to work with communications professionals to effectively and transparently communicate the state's assessment plans, especially if plans involve changes from previous practices.

[1] 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. Center for Assessment (www.nciea.org) and The Aspen Institute (www.aspeninstitute.org/education).

[2] Marion, S. F. (2020). Using Opportunity-to-Learn Data to Support Educational Equity. Center for Assessment. Retrieved from: https://www.nciea.org/articles/using-opportunity-learn-data-support-educational-equity

IN MARCH OF 2020,

the threat of the COVID-19 pandemic reached a tipping point in the United States, disrupting nearly every aspect of daily life. This included the K-12 educational experience: School buildings quickly closed and transitioned to remote models, and our collective attention rightly focused on supporting the fundamental needs of students and schools facing unprecedented challenges.

Large-scale assessment administration essentially became a non-issue in spring 2020. Indeed, before March turned to April, the U.S. Department of Education (ED) granted waivers to all 50 states, D.C., Puerto Rico, and the Bureau of Indian Education to suspend 2020 testing and accountability.

Many schools expected to return to typical operations in fall 2020 and planned to resume large-scale assessments in spring 2021. However, the pandemic did not subside, and most schools started fall 2020 with remote or hybrid learning. "As of September 2," the 9/21/20 issue of Education Week reported, "73% of the 100 largest school districts have chosen remote learning as their primary back-to-school instructional model, affecting over 8 million students."

Given the state of the states in fall 2020, the prospect of moving forward with "normal" state summative testing in spring 2021 is increasingly uncertain. Unlike in spring 2020, however, there is less agreement about the best path forward. State leaders remain committed to having a clear sense of what students in their states are learning to better understand what support students and schools may need. How state leaders are working to achieve that goal, however, varies based on a range of considerations, including uncertainty about the operating status of schools in their states by the Spring, among other things. While some leaders call for suspending or relaxing assessment requirements in 2021, others are striving to administer the state summative tests in the Spring. In September, U.S. Education Secretary Betsy DeVos sent a letter to chief state school officers, conveying that "states will, in the interest of students, administer summative assessments during the 2020-2021 school year, consistent with the requirements of the law." Acknowledging that this year's assessments may "look different," the Secretary urged state education leaders to demonstrate resourcefulness in identifying workable solutions. Nevertheless, some states' resourceful solutions may face approval challenges if their proposals are perceived as straying too far from the law.

CCSSO has stated that it believes in the importance of high-quality assessments and that measuring student learning and identifying potential gaps is more important than ever, given the current environment. In its most recent statement released in December 2020, CCSSO underscores that high-quality assessments are a crucial way to measure student learning, identify inequities, and drive the right supports to students. Given that the impact of COVID-19 varies across the country, CCSSO notes that "states are moving forward and exploring how to administer their statewide summative assessment to as many students as possible this spring, or exploring other similar important measures."

When it comes to states administering summative tests in spring 2021, there are challenging questions to be sure, they have no easy answers, and it is increasingly apparent that any proposed solutions must include some unconventional alternatives as back-up plans.

UNDERSTANDING THE CHALLENGES TO STATE TESTING

One must understand the problem before identifying a credible solution. While some of the challenges outlined below are likely to be common to all states, others may be associated with a particular set of values, priorities, or circumstances. For example, one state may view the primary challenge as lost instructional time. In this circumstance, any solution failing to address efficiency (e.g., minimizing the time burden on teachers and students) may be regarded as insufficient. Another state may see the primary challenge as the inability to test students in school buildings. Here, a workable solution must address test administration alternatives, such as expanding the testing window or providing remote administration. Some states doubtless will be in need of solutions that overcome multiple obstacles. In any case, it is important to be clear about the connections between proposed solutions and the challenges, which can be grouped into three categories: EQUITY, UTILITY, and CAPACITY.

EQUITY

Inequity has been an ever-present and institutionally intractable problem in U.S. public education. Of particular concern is that the disruptions caused by COVID-19 will only amplify inequities for those students who already are facing inequities in remote or hybrid learning situations.

The digital learning gap is real. Not all students have adequate or reliable access to the Internet, much less the proper electronic devices,—to meaningfully engage in remote learning. Even if there is a computer at home, bandwidth may be inadequate and/or screen time may be limited, because one electronic device must be shared among multiple family members. News reports of students seeking a Wi-Fi connection in a fast-food parking lot offer chilling testimony to the differential access to resources for public school students in America (Harris, 2020). This not only hampers a student's ability to participate in online assessments; more importantly, it hampers the student's opportunity to learn.

Moreover, learning conditions at home can be highly variable. Some students have a quiet, well-equipped work environment that accommodates the demands of remote learning and assessment; others do not. Some students further benefit from home enrichment, such as having family members who provide instructional support, while others do not. Again, such inequities have always existed, but they are more extreme and more consequential in the remote or hybrid learning environment. As Levine (2020) observes, educators' efforts to promote equity within the walls of a brick-and-mortar school – such as ensuring equitable access to instruction, equipment, and materials – are so severely threatened in a remote environment that "total amount of instructional time is likely to plummet."

Finally, many learners need specific supports that schools are distinctly equipped to provide in person. During in-person instruction, English language learners and students with disabilities receive specialized instruction, targeted adaptations, and unique resources – and such accommodations and modifications are not generally possible to replicate with fidelity in a remote-learning environment. Even in a hybrid-learning environment, there may not be enough in-person time available to provide this important, additional support.

Each of these issues likely will impede the opportunity to learn and prevent meaningful participation in assessments, especially for the most vulnerable students. The gravity of these equity issues cannot be overstated. It is a moral imperative to evaluate any proposed assessment solution with this criterion in mind:

Does it meet the needs of <u>all</u> students?

UNDERSTANDING THE CHALLENGES TO STATE TESTING

UTILITY

Given the COVID-19 disruptions to in-person learning that began in spring 2020, it is reasonable to inquire whether the results from large-scale state assessments in 2021 can be used meaningfully. One concern is that the content represented on the state test will not meaningfully reflect the instructional experience of students. Existing test blueprints may be based on standards that were skipped, or covered insufficiently, in the midst of school building closures and curricular shifts necessitated by remote and hybrid models of learning. Further, the variability in school contexts may create a situation in which a "one size fits all" state test is especially problematic.

Even where test blueprints have been adjusted, this may not quell concerns about uneven opportunities to learn—due, in part, to the aforementioned equity issues. In addition, there are concerns that new models for teaching and learning are being rolled out before they can be fully developed. As such, it is not surprising that some researchers have reported tremendous variability among teaching-learning models with regard to the quality of implementation, the degree to which students are engaged, and the level of support being provided for students with special needs (Barnum & Bryan, 2020).

Given these concerns, stakeholders understandably may doubt whether test scores can be meaningfully interpreted and used. Such doubts primarily reflect the ways in which many stakeholders conflate assessment and accountability—a common enough viewpoint, since test scores play a central role in supporting inferences about the performance of educators and schools. The practice of using test results as a basis for attribution (e.g., taking low proficiency rates to signal that a school is poorly serving its students) is controversial in the best of circumstances. This controversy is only heightened in situations where so many external factors are inhibiting the opportunity to learn.

CAPACITY

This paper uses the term *capacity* to describe two pragmatic threats. The first pertains to logistics: Education leaders may face the realization that they cannot safely test students in school buildings consistent with previous practice. Testing in school settings, under the direction of a well-trained examiner, helps to ensure that assessment content remains secure, administration conditions are standardized, and students are given appropriate accommodations and supports. If testing cannot occur in schools, the meaning and comparability of the test results may be questionable.

Secondly, a state's policy priorities may lead to other capacity issues. That is, some education leaders may be reluctant to use valuable resources to support state summative assessments, particularly in light of the many threats considered above. Indeed, standardized testing expends two of the most precious resources in education: time and money. Instructional time was already at a premium before the pandemic, and it is even more so when schools are coping with disruptions and changes. Moreover, with the nation's current economic downturn reducing already-slim education budgets, state leaders understandably will be seeking ways to recover funds—such as those used to support state testing.

EXPLORING SOLUTIONS

Below are several alternatives to status-quo testing, some of which may be regarded as unconventional. Indeed, many options have known drawbacks. Others have not yet been implemented at scale, and this is not a trivial concern: the practice of large-scale assessment rests heavily on established practices to bolster the reliability, validity, and fairness of score interpretation and use.

Moreover, the authors acknowledge that some of the options presented in this paper are not consistent with the statutory requirements of the Every Student Succeeds Act (ESSA). To be clear, an option will not meet federal requirements if it does not produce assessment results for all students (a) in reading/language arts and mathematics, once in each of grades three through eight and once in high school; and (b) in science, once for each required grade band. Additionally, assessments must meet requirements for validity, reliability, and comparability to comply with ESSA.

Some of the options below are clearly outside these parameters (such as sampling students, since it would eliminate some tests). Other options may pose a risk to compliance (such as adjusting the test blueprint or changing test administration conditions). In such cases, the impact of consequences will depend on the nature and extent of the changes. As Secretary DeVos's letter invites, the authors' exploration of alternatives was unconstrained by current federal policy, which permitted consideration of a broader range of possibilities.

The following four considerations, and the primary question each one addresses, serve to frame this paper's proposed alternatives:

ACCESS & SUPPORT

How can states bolster the capacity of schools and districts to address the needs of all students, particularly those who may be most vulnerable? While this is not a testing alternative, it is a critical overarching concern.

DESIGN

How might states select or revise the assessment instrument?

ADMINISTRATION

How can states modify the manner in which the assessment is implemented?

INTERPRETATION & USE

How should states help stakeholders understand and use the results in an appropriate and defensible manner?

These connections are not one-to-one; rather, some alternatives may address multiple threats. While the alternatives are not mutually exclusive (i.e., implementing one option does not preclude implementing another) there may be diminishing value in combining multiple alternatives associated with design and administration. Ultimately, there are benefits to identifying the most direct path forward: one that is no more complicated and onerous than it must be.

EXPLORING SOLUTIONS: ACCESS & SUPPORT

ACCESS & SUPPORT

While technically not an assessment alternative, any initiative to increase access and support should help states implement the observations that follow.

All students should have sufficient access to the resources necessary for full participation in learning and assessment. This includes devices—computer, software, ancillary equipment—and adequate Internet connectivity. It also includes assistive technology as needed, such as text-to-speech software and other supports for students with disabilities.

How can states accomplish this? It begins by conducting a comprehensive needs assessment similar to suggestions for collecting and reporting opportunity-to-learn information.[3,4] State leaders can work with local education agencies to provide guidance and bolster capacity to identify and collect information about the availability of resources for learning and assessment. It also will be important for data collection to include check-ins to gauge the extent to which students with disabilities and English language learners have access to the supports and services they need, especially in remote or hybrid-learning situations.

If the state is considering remote assessment alternatives, it will be essential to conduct an early evaluation of readiness and capacity in this regard. For example, many states and service providers have already developed guidance to help "traditional" testing sites evaluate readiness for assessment, such as by specifying minimum technology requirements and outlining procedures to check that hardware and software systems are properly configured and sufficiently robust to support administration of the assessment. These requirements and procedures should be reviewed and adapted as necessary—well in advance of the testing window.

To be sure that all will proceed smoothly on the day of the test, gaps in access to necessary resources should be addressed as early as possible. CARES Act funding may be available to support some needs; in other cases, states might provide grants or develop partnerships with organizations that could help districts and schools meet student needs.

There are no easy solutions to social inequities that are stark, complex, and persistent. That said, equitable access to needed electronic resources is critical for all students' participation in remote learning. If states leave these inequities unaddressed, virtually none of the assessment options discussed below will be viable.

[3] 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. Center for Assessment (www.nciea.org) and The Aspen Institute (www.aspeninstitute.org/education).

[4] Marion, S. F. (2020). Using Opportunity-to-Learn Data to Support Educational Equity. Center for Assessment. Retrieved from: https://www.nciea.org/articles/using-opportunity-learn-data-support-educational-equity

EXPLORING SOLUTIONS: DESIGN

DESIGN

Presented here are three options for selecting or revising an assessment. Note that the technical aspects of implementing these options are described in greater detail in Keng, Boyer, and Marion's 2020 paper, *Into the unknown: Assessment considerations for spring 2021*.[5] These options are summarized and compared in Table 1 (below).

REUSE A PRIOR FORM

One alternative — reusing a prior test form — is not a particularly uncommon practice, which is a major benefit. A decided advantage of this option is that a previously administered form will have both an established score scale and stable item parameters. It is a daunting technical challenge to create a trustworthy and comparable new scale in a year when, to say the least, learning and administration conditions may be uneven (Keng et al., 2020). Reusing an existing form, therefore, will better position the state to measure changes in student achievement from 2019 to 2021. Moreover, this practice addresses the capacity threat, insofar as it is markedly less expensive to recycle a previous form than to develop a new one. The resulting savings could be used to support other priorities.

The primary drawback of this approach is security, particularly if there is only one fixed form (as opposed to an adaptive test) and it contains performance tasks or "memorable" items. If content from a prior administration has been circulated, student performance outcomes may be inflated. Nevertheless, the authors believe that the benefits of reusing a prior form outweigh the risks, particularly for a one-off administration.

ADJUST BLUEPRINT

Another design alternative is to alter the test blueprint. This option may be appealing for at least two reasons: First, if the state has identified a subset of prioritized standards (e.g., Student Achievement Partners, 2020) to mitigate opportunity-to-learn threats, adjusting the blueprint to reflect that content produces a measure that is likely to be more consistent with what teachers should be teaching and students should be learning. This improves the usefulness of test results.

A shortened test is the second reason to adjust the blueprint. For example, a state may decide to reduce the number of items on the form to decrease testing time. Removing performance tasks (such as writing prompts) or eliminating items associated with lengthy stimuli (such as reading passages) would be particularly effective in reducing testing time. A shorter test will recover precious instructional time as well as accommodate more frequent test administrations with fewer students present at each one.

The primary limitation of adjusting the test blueprint is that it may impede comparability of the results across successive-year administrations of the test. For example, if 2019 test content (and therefore the corresponding blueprint) is changed in 2021 to reflect a subset of prioritized standards, it may be impossible to determine whether differences in performance from 2019 to 2021 reflect a true shift in student achievement or, rather, a difference in performance expectations between the two forms. However, this drawback is not unique to this approach; even if "regular" testing is maintained, it still will be challenging to disentangle shifts in achievement from changes in opportunity to learn.

In contrast, it may be possible to preserve comparability of results if the test length is reduced but content remains generally representative of the existing blueprint. In this instance, the chief threat shifts to the reliability or precision of the scores. That is, are the scores based on the streamlined test as trustworthy as those from the longer test? In most cases, the effect of form reduction on reliability can be estimated in advance of test administration. We encourage states to work with their Technical Advisory Committee to determine the best way to ensure sufficient reliability for the test to support its intended use. For example, a substantial reduction in test length may be defensible for states that conventionally report fewer subscores or performance levels.

[5] 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

EXPLORING SOLUTIONS: DESIGN

SAMPLING

Sampling encompasses various methods for measuring a subset of a targeted group or population. In the context of test design, this might involve sampling students, test items, or a combination of the two.

Item (Matrix) Sampling

Item or matrix sampling is technically a type of design change related to the blueprint, but the authors believe it is sufficiently distinctive to merit separate description. Item, or matrix, sampling shortens an assessment for a student by distributing items across multiple students. That is, each student takes only part of the whole form. When these multiple parts are combined, information can be reported summarily at the district, school, and student-group level. There are many types of matrix designs, some of which preserve at least partial data at the student level. More typically, however, matrix designs only support summary reporting. By substantially reducing the test form for any one student, matrix sampling can recover instructional time as well as enable schools to offer smaller (and thus safer) in-person administrations.

One drawback of matrix sampling is that it may preclude reporting student-level scores (either partially or fully), although this may be regarded as an advantage by those states that are concerned about the misuse of student-level results. Another drawback is that matrix sampling can be complicated to design and implement; if a state is not already planning for the spring 2021 test administration by the time of this paper's publication, it will be a tall order to implement a suitable matrix-sampling plan with fidelity in time for spring 2021 testing. Moreover, for the more than 25 states that employ a computer-adaptive testing (CAT), conventional matrix sampling does not apply. In some ways, CAT can be regarded as an extreme example of matrix-sampling, but the state can still choose to shorten the test to accomplish a goal similar to matrix sampling.

Sampling Students

Sampling students refers to strategies to reduce the number of assessments administered. As described above, this may be appealing to districts that wish to free up resources (such as instructional time and funds) to use for other priorities. It may also help districts and schools seeking to administer tests concurrently to fewer students.

One way to reduce the number of administered tests is to transition to grade-band testing state-wide. For example, instead of testing ELA and mathematics in each of grades 3-8, test only in grades 5 and 8. This would substantially reduce the number of tests given in elementary and middle schools, while still providing some information for this grade band. This approach has the benefit of allowing for socially-distant testing, within a reasonable testing window. Moreover, removing a test entirely results in the most cost-savings. The obvious drawback of sampling students in this fashion is the considerable loss of information on student achievement. Not only does it preclude trend comparisons for grades no longer tested, it also complicates — and possibly prohibits — calculations of student growth.

Another approach is sampling students within grade. For example, students could be assigned to take either the ELA or the mathematics test in each of grades 3-8. If student assignment to testing condition were based on an appropriate sampling frame, this approach preserves the availability of aggregate scores for each grade and content area, while markedly reducing overall testing. That said, cost savings is less than it is if one or more tests are eliminated. Within-grade sampling also adversely affect calculations of student growth. Finally, the within-grade sampling of students could be particularly challenging, because parents may refuse to have their child participate in in-person testing. If this occurs, the reduced levels of participation could compromise the representativeness of the sample, generalizability of test scores, and validity of the inferences stakeholders ultimately wish to make. For all of these reasons, within-grade sampling may not be a viable option.

EXPLORING SOLUTIONS: DESIGN

TABLE 1 STATE SUMMATIVE ASSESSMENT DESIGN ALTERNATIVES FOR 2020-2021

OPTION	BENEFIT(S)	LIMITATION(S)
REUSE PRIOR FORM	· saves time and money, which can be used to support equity priorities · bolsters comparability	· potential security threat
ADJUST BLUEPRINT – PRIORITIZED STANDARDS	· may reduce test length, recovering more time for instruction · may strengthen the link between instruction and assessment	· inhibits trend comparability · difficult to implement in a short time frame
ADJUST BLUEPRINT – SHORTEN THE TEST ONLY	· reduces test length, recovering more time for instruction	· may reduce score precision · may necessitate loss of test information (e.g., subscores)
ITEM SAMPLING	· reduces test length, recovering more time for instruction	· likely permits only aggregate scores · difficult to implement in a short time frame
SAMPLING STUDENTS	· fewer groups of students concurrently testing mitigates safety threats	· loss of test information at the student level · hinders calculation of academic growth in the near and longer term

EXPLORING SOLUTIONS: ADMINISTRATION

ADMINISTRATION

Administration alternatives reflect which tests are administered, and how. These options, summarized in Table 2 (below), are particularly beneficial for addressing capacity-related challenges.

Expand the Testing Window

We begin with arguably the least disruptive alternative to administration: increasing or lengthening the testing window. This involves extending either the time frame in which the state allows districts to select test administration dates or, more directly, the length of the window within which districts or schools are permitted to administer the test. Certainly, both timing and duration also can be modified to offer maximum flexibility.

Most states currently have a fairly large window within which to determine the timing and duration of testing. Nevertheless, expanding the testing window permits more flexibility for "cycling" students through the test administration, which mitigates safety concerns by testing smaller, more spread-out groups of students. If combined with test-length reductions, this may further bolster capacity for small-group testing.

There are, however, at least two concerns associated with this alternative. First, any time the test window is expanded, there is an increased risk that test content will be shared. However, many assessment leaders likely will conclude that this security risk is significantly offset by increased safety benefits — particularly where assessment has been decoupled from accountability.

A second concern is differential opportunity to learn. If the testing window is expanded too much, there will be a pronounced—and inequitable—difference in opportunity to learn for early testers, as compared with later testers. In short, the window should not be so broad that students testing early are denied adequate instruction of material that is included on the exam.

Remote Administration and Proctoring

Interest in remote administration has been kindled by the likelihood that safe in-person testing may not be possible in spring 2021. Therefore, states may well consider allowing students to take the test in a remote location, presumably the student's home.

The obvious challenge for a remote administration is to maintain the integrity of the assessment experience when it is not directly overseen by a well-trained administrator. Short of implementing a proctoring solution, which we address next, there are various strategies to increase the likelihood that the administration is implemented with fidelity. They include:

- Produce and circulate a testing 'checklist' that clarifies requirements for a successful remote administration (e.g., outlining technology requirements, describing intended testing conditions, etc.)
- Survey students and families to determine the extent to which the requirements can be met. Work with families to provide support as needed. This is particularly important for students with special needs and English language learners who may require individual supports that may not be available outside of school.
- Provide opportunities for training and practice sessions to ensure that students are familiar with the testing platform and it is working as intended.
- Specify procedures for contingencies (e.g., What should you do if the system goes down during test administration?)
- Circulate a post-administration survey to collect information about the testing conditions and experience. This data can be used to help states determine the extent to which the test results will be useful for their intended purpose.

EXPLORING SOLUTIONS: ADMINISTRATION

Remote Administration and Proctoring (cont.)

Naturally, standardization and security will be difficult, if not impossible, to maintain in a remote setting. For example, features that work well in a controlled, proctored setting, such as prohibiting users from accessing the Internet during the test, can be easily circumvented by using another device or seeking assistance from other people or resources.

Consequently, some states may want to explore remote proctoring as a means to bolster the integrity of a test that is not administered in a central location. Remote proctoring is not new; it has been implemented for many years, mostly for online professional licensure and certification assessments. It appears, however, that remote proctoring has not, to date, been employed at scale for K-12 state summative testing.

There are three approaches to remote proctoring: human-based, technology-based, and hybrid. A human-based approach brings a test proctor and examinees together online, and the proctor monitors test administration via microphone and Webcam. The examinee-to-proctor ratio, of course, must be low enough to be practical and permit successful monitoring. In contrast, a technology-based approach uses software to do the monitoring. Here, too, examinees establish an online connection involving Webcam and microphone. During testing, the proctoring software detects irregular noises and movements that are associated with testing anomalies and, after testing, reports the likelihood that such anomalies have occurred. Users should carefully consider the evidence associated with any artificial-intelligence solution, given the known limitations (e.g., racial bias). Finally, a hybrid approach combines the two.

Overall, there are substantial challenges associated with remote administration. States considering this option should collect evidence addressing four primary threats (Keng et al., 2020).

  COMPARABILITY: Are the scores obtained from remote administration comparable to those from standard test administrations?

  TECHNOLOGICAL ACCESSIBILITY: Do all students have sufficient technological capacity to participate as intended?

  ACCOMMODATION REQUIREMENTS associated with the Individuals with Disabilities Education Act and Americans with Disabilities Act Accessibility: Do students with disabilities have the support they would receive with conventional administration?

  SECURITY: Are sufficient safeguards in place to prevent improprieties?

Two additional concerns deserve mention. First, will remote test administration receive public support, given the novelty of remote proctoring in K-12 grades and likely privacy concerns? Parents' objection to student participation in state standardized testing is not new, and this resistance may be intensified with an approach regarded as invasive. Second, students' motivation is likely to differ when they are taking a test under normal conditions, as compared with taking a test remotely, and these differences cannot be ignored.

We acknowledge the appeal of continuing the state test without a standard in-school administration. Nevertheless, if the goal is to produce standardized, comparable scores, the authors believe the potential threats of remote administration with or without proctoring are sufficient to question its feasibility in the near term.

EXPLORING SOLUTIONS: ADMINISTRATION

DISTRIBUTED ASSESSMENT

Another alternative is to transfer primary responsibility for test administration from the state to the local education agency (LEA). The distributed-assessment alternative may be appealing to states that conclude a "one size fits all" approach to assessments is not feasible for districts operating under widely varying learning models (fully remote, in-person, or hybrid).

While an exhaustive treatment of the distributed-assessment alternative is beyond the scope of this paper, two options merit consideration. The first option is to decentralize administration of the **standard state test**, with guidance for local administration. The guidance would be far less restrictive than it would be for a standard administration, allowing LEAs broad flexibility for students to take tests either remotely or in-person, as circumstances warrant. Paper-based tests could be scored locally (e.g., with state-provided score tables) or returned for central scoring. In this scenario, the state is essentially yielding on security and standardization requirements in order to promote the flexibility required to make testing both safe and accessible. The shared understanding, in this case, is that the materials provided to the LEA are nonsecure following the event; moreover, it is understood that test results will not be comparable to data derived from previous assessments—nor will it be used for high-stakes purposes.

A second option is for the state to provide LEAs with guidance for administering a **locally developed assessment** that meets state criteria. For example, the state may establish criteria for assessment content that would be developed, administered, and scored at the district or school level. The assessment could be similar to the traditional state test or take the form of a portfolio, a project, performance tasks, etc. This model positions the state to serve more as a partner helping LEAs to implement strong practices to evaluate the extent to which their students are meeting state expectations. Again, in this scenario the test results will not be regarded as standardized, comparable, or suitable for high-stakes purposes.

TABLE 2 STATE SUMMATIVE ASSESSMENT ADMINISTRATION ALTERNATIVES FOR 2020-2021		
OPTION	**BENEFIT(S)**	**LIMITATION(S)**
EXPAND TESTING WINDOW	• fewer groups of students concurrently testing mitigates safety threats	• potential security threat • could impede OTL for some students
REMOTE ADMINISTRATION (PROCTORING)	• maintain testing program without the safety risk associated with in-person administration	• comparability of scores is not assured • access to technology and supports for students with special needs must be addressed • additional cost • privacy concerns may prompt opt-outs
DISTRIBUTED ASSESSMENT	• shifts focus to helping LEAs and schools develop and implement flexible practices to gauge student achievement	• results are not comparable or suitable for interpretation and use in a similar manner • differences in conditions and capacity may lead to uneven implementation

EXPLORING SOLUTIONS: INTERPRETATION & USE

INTERPRETATION & USE

Although this paper has already addressed elements of interpretation and use, the authors believe this issue is sufficiently important to elaborate upon further. As noted earlier, using summative assessments in spring 2021 as part of a state's status quo accountability system is likely to be problematic. If a state chooses to make changes to its accountability system, some form of approval from ED will be required. Given the impediments to adequate opportunity to learn, as well as the difficulty in producing scores that can be meaningfully compared with prior years, states should consider decoupling spring 2021 assessments from high-stakes accountability for students, educators, and schools.

How, then, should 2021 test results be interpreted and used? Under most conditions described in this paper, public reporting would be justifiable *if* accompanied by strategic messaging to mitigate the risk of data being misinterpreted and/or misused. States might consider revising reports by providing additional narrative and visual cues to guard against unwarranted interpretations (Domaleski, Boyer, & Evans, 2020). Where the assessment blueprint has changed, for example, or where administration conditions are not standardized, trend data may be compromised. Additionally, it will be important for states to develop customized communication supports, such as interpretation guides and training materials, to help stakeholders understand and use assessment information appropriately.

Even if assessment results are decoupled from conventional accountability, the data can support a state's school-improvement initiatives. We suggest exploring what some have termed "little a" accountability, which is shorthand for initiatives that fall outside the accountability system developed to meet ESSA requirements. For example, states and districts may use assessment results, along with other indicators, to inform how to modify improvement plans, deploy resources, and establish professional development plans. Finally, states could design and implement an opportunity-to-learn data collection and indicator system, to contextualize and interpret test scores from 2021 (Marion, 2020, Marion, et al., 2020). Even if test data is not available, collecting opportunity-to-learn data could provide important insights with regard to allocating future resources to the schools and students most in need of support (see Table 3 below).

TABLE 3 STATE SUMMATIVE ASSESSMENT INTERPRETATION & USE ALTERNATIVES FOR 2020-2021		
OPTION	**BENEFIT(S)**	**LIMITATION(S)**
DECOUPLE ASSESSMENT FROM HIGH-STAKES ACCOUNTABILITY	· mitigates risk of unintended negative consequences	· may require development of modified reporting system and/or revision of school support strategies
REPORT ASSESSMENT RESULTS & INTERPRETATIVE INFORMATION IN THE CONTEXT OF PRIORITIZED OTL INFORMATION	· helps users understand certain influential factors on test scores and will hopefully promote more honest interpretations	· requires being able to collect high-quality OTL data

PLAN & ACT NOW

State testing and other data collection efforts (e.g., OTL) operate on lengthy development and quality-control cycles. Trying to compress this time cycle too aggressively could result in costly and embarrassing mistakes. State leaders are encouraged to work closely with their assessment experts and technical advisors to begin planning now for multiple contingencies later in this school year. It will be important for state leaders and assessment specialists to work with key stakeholders (e.g., district leaders, critical policymakers) to incorporate their input and vet any plans for altering assessments. It will be especially vital to evaluate any proposed assessment alternatives in light of potential challenges to equity, utility, and capacity — as well as other areas the state itself might identify. In reviewing proposals, state leaders will want to prioritize solutions that advance equity and access for all students. Finally, and perhaps most importantly, once they decide on a course of action state leaders will need to work with communications professionals to effectively and transparently publicize the state's assessment plans— especially if those plans involve making changes to previous practices.

When it comes to states administering summative tests in spring 2021, there are challenging questions to be sure, they have no easy answers, and it is increasingly apparent that any proposed solutions must include some unconventional alternatives as back-up plans.

REFERENCES

Barnum, M. & Bryan, C. (2020, June 26). America's great remote-learning experiment: What surveys of teachers and parents tell us about how it went. Chalkbeat: Teaching & Classroom. Retrieved from: https://www.chalkbeat.org/2020/6/26/21304405/surveys-remote-learning-coronavirus-success-failure-teachers-parents

Domaleski, C., Boyer, M., & Evans, C. (2020). Accountability interrupted: Guidance for collecting, evaluating, and reporting data in 2020-2021. Council of Chief State School Officers: Washington, D.C. https://ccsso.org/sites/default/files/2020-09/CCSSO_RR_Accountability-v3.pdf

Harris, B. (2020). Homework in a McDonalds parking lot: Inside one mother's fight to help her kids get an education during coronavirus. The Hechinger Report. Retrieved from: https://hechingerreport.org/homework-in-a-mcdonalds-parking-lot-inside-one-mothers-fight-to-help-her-kids-get-an-education-during-coronavirus/

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

Levine, Peter (2020, May 14). Educational equity during a pandemic. Albert Shanker Institute. Retrieved from: https://www.shankerinstitute.org/blog/educational-equity-during-pandemic

Marion, S. F. (2020). Using Opportunity-to-Learn Data to Support Educational Equity. Center for Assessment. Retrieved from: https://www.nciea.org/articles/using-opportunity-learn-data-support-educational-equity

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*. Center for Assessment and The Aspen Institute. Retrieved from: https://www.nciea.org/articles/not-test-emergency-special-considerations-assessing-and-advancing-equity-school-year-2020

School districts' reopening plans: A Snapshot (2020, July 15). Education Week. Retrieved from: https://www.edweek.org/ew/section/multimedia/school-districts-reopening-plans-a-snapshot.html

Student Achievement Partners (2020). 2020-2021 Prioritized Instructional Content in English Language Arts/Literacy and Mathematics. Retrieved from: https://achievethecore.org/page/3267/2020-21-priority-instructional-content-in-english-language-arts-literacy-and-mathematics