Assessment and Accountability in the Wake of COVID-19

Introduction

As states and school districts closed schools in the midst of the COVID-19 virus outbreak, nearly 55 million American K-12 students were abruptly sent home and forced to continue their education remotely. Within days, the U.S. Department of Education announced flexibility for states to cancel their annual summative assessments and school accountability ratings for the spring of 2020. Every state applied for and received the waivers by the end of the month.¹ Waiving federal testing requirements and postponing accountability decisions was the right policy move at the time, for a number of reasons. The logistics of operating schools safely in the midst of a global health pandemic were then and are still likely to be immense.

As school districts are busy planning for what’s likely to be a tumultuous 2020-21 school year, it might be tempting to continue the testing and accountability pause for another year. That would be a mistake. Instead, state leaders should be planning now to make their testing and accountability systems resilient to changing circumstances on the ground. Another year of waivers would represent a failure of planning and imagination that would reinforce and amplify existing inequities in our school system. The transition to remote learning has gone poorly in many places, and the consequences have disproportionately fallen on low-income students, English learners, students with disabilities, and homeless students. States must find a way to assess student learning gains — or losses — and target support to the students and schools that need them the most.

While testing data is the most obvious way to collect information on student learning, states are also missing a number of other important measures of quality and equity across our schools. States will need new ways to collect data on student attendance, engagement, high school graduation rates, college and career readiness, and other input and outcome indicators.
Ultimately, school districts are “creatures of the state.” Districts were created by state legislatures and exist to carry out the state’s constitutional obligation to provide a public education to all students. It is state leaders who define the minimal requirements for a school day and year, set grade-level standards for what students should know and be able to do, and grant high school diplomas to students who have met the state’s requirements. To carry out these responsibilities well, state leaders must be thinking now about the adjustments they’ll need to make to restart their testing and accountability systems.

This brief begins by describing the need for states to monitor and track student performance in the wake of the COVID-19-related school closures. It continues by outlining some of the challenges states will face to restart their testing and accountability systems, as well as potential solutions.

Why Accountability Systems Are Even More Important Now

As American school systems shut down their schools in response to the coronavirus pandemic, they were asked to deliver instruction from a distance. Although the specific mechanisms varied widely, districts asked teachers to offer live online instruction, mailed or emailed packets of information to families, or used other delivery mechanisms like public broadcasting stations.

There were good reasons from the start to doubt whether this transition would go well, particularly for the youngest and most disadvantaged students. Research on schools and courses that have been explicitly designed to be delivered from a distance finds that those classes tend to be less effective than courses delivered in-person. Moreover, the students participating in those studies tend to be adults, or at least older K-12 students who have voluntarily chosen to take virtual classes. In those settings, instructors have time to plan and prepare for the logistical and pedagogical demands of delivering instruction virtually. In contrast, in the U.S. alone, the coronavirus abruptly forced 13,000 school districts, with 100,000 schools, 3.5 million teachers, and 54 million students from ages 5-18, into a virtual learning environment involuntarily.

It would be bad enough if the COVID-19 school closures hit all students evenly. But that’s not what happened. Students from wealthier families were more likely to have access to internet-connected devices, to have high-speed internet at home, and to have other educational materials in their home. These advantages helped buffer wealthier students from the lost learning time and allowed them to stay engaged with distance learning.

School district responses also varied, and those responses correlated with family income. For example, the Center for Reinventing Public Education found that less than half of school districts took attendance or required teachers to check in with their students, and affluent districts were twice as likely as
high-poverty districts to offer live instruction. According to a nationally representative sample from the University of Southern California, about 63 percent of families earning less than $25,000 a year reported that their child had interacted with a teacher since the shutdowns began, compared to 88 percent for families earning $150,000 or more. Teachers report similar trends: According to surveys conducted by the EdWeek Research Center, teachers estimated they were delivering three hours of instruction per day, a decline of three hours from prior to the shutdowns. But for students in high-poverty districts, teachers estimated they were delivering just two hours of instruction.

The shutdowns affected students in many ways, and we may never be able to quantify the full effects in terms of student mental health or emotional trauma. We will, however, eventually have more data on academic learning losses as states and districts restart their assessment systems. For now, we can only guess at the size of those effects. When researchers for NWEA, the Northwest Evaluation Association, analyzed historical data from more than 5 million students who participated in NWEA's widely used MAP test in 2017-18, they projected that students could return next fall with as little as 50-70 percent of the math and reading gains they would have made in a normal academic year.

While we will start to see hard data on the student learning losses when they come back in the fall, it's likely that the full impact won't be known for years. For now, we can guess at those effects by looking at the research on past school closures. That body of research finds negative effects on two levels.

First, school closures can disrupt educational trajectories if students use the opportunity to make different life decisions than they otherwise would have. This will mostly affect older students, in high school or college, who are at risk of dropping out now in order to take advantage of immediate job opportunities. In 1916, for example, schools in many places closed for several weeks in the midst of a polio outbreak. Working-age students who lived in communities with more cases of polio eventually completed fewer years of schooling.

Second, researchers have attempted to quantify the extent of the learning loss from school closures in a number of contexts. Natural disasters provide one parallel. For example, test scores fell in Thailand after severe flooding in 2011 forced schools in some provinces to close for up to a month. A study looking at the aftermath of a 2005 Pakistan earthquake found that the children of well-educated mothers suffered no educational impacts at all, while other children showed substantial losses. Worse, the test score losses grew over time, even once kids were back in school.

A series of teacher strikes in Argentina provide an alarming example of how lost learning time affected the economic outcomes of students and their eventual children. Starting in the early 1980s, Argentinian provinces suffered through a wave of teacher strikes that caused primary school students to miss an average of 88 days of learning, or about 7 percent of their total primary school days. A research study looking at the long-term impacts of the strikes found that, for young students exposed to the lost learning time, the strikes caused a decline in high school and college completion rates, fewer years of education completed, and lower employment rates and earnings. The authors estimated that 30- to 40-year-olds who were school-aged during the strikes earned 3 percent less than they otherwise
would have, which amounted to an annual loss of more than $700 million to the Argentine economy. The study also found that the negative effects passed on across generations. Once the strike-affected children grew up and had children of their own, the children of the strike-affected children were more likely to be held back in school compared to their peers.

It’s possible that these studies are too pessimistic for the current moment. After all, they all existed in eras and locations without any real attempt at distance learning. Plus, the COVID-19 school closures occurred during the final months of the school year. In mid-March when schools began shutting down, many districts had their scheduled spring break weeks, not to mention the spring testing window that can take the place of ordinary instructional time. In general, the last few weeks of school are not known as the most rigorous ones of the year, as teachers and students begin looking forward to summer.

But the research suggests that school closures affect different students differently. As might be expected, the learning losses during school closures are the most pronounced for younger students and for disadvantaged students. Because younger students learn much more over the course of a year than older students do, they also have more to lose by not being in school. Younger students suffer larger summer learning losses compared to older students, and younger students suffered larger and more persistent negative effects in the aftermath of the Thailand floods and the Argentina teacher strikes. And when researchers Dave Marcotte and Steven Hemelt looked at the effects of snow days on Maryland children, they found larger learning losses for third-graders than for fifth- or eighth-graders.

There are other signs that these divergences are playing out here in the U.S. as well. According to data from Harvard University’s Opportunity Insights Economic Tracker, students in high-income ZIP codes were making more progress in online math courses during school closures than they had been in prior months. In contrast, students from low-income ZIP codes faced declines of 25-35 percent during the shutdowns.

Given this body of historical evidence, it seems likely that the COVID-19 school closures are likely to cause significant learning losses, especially for the youngest and most disadvantaged students. The American education system has never been a paragon of equity, but the COVID-19 closures are making those problems much worse. We won’t truly know the full effect, or how it differs across schools, districts, and different populations of students, until states resume testing all students. Well-designed accountability systems can help identify and target resources to the students who will need the most help. Without action from state leaders, or with only universal, one-size-fits-all responses, achievement gaps are likely to expand significantly.

It seems likely that the COVID-19 school closures are likely to cause significant learning losses, especially for the youngest and most disadvantaged students.
Questions for States as They Rebuild Their Accountability Systems

As state leaders plan for the 2020-21 school year, it seems likely that some or all school districts in their states will be hit by further COVID-19 outbreaks. Even if schools somehow manage to have a normal school year, states will still need to adjust accountability systems for the missing data from 2020. This section outlines some key questions state leaders will need to face as they move forward.

Will states change how many and which schools are identified for comprehensive and targeted support?

Under No Child Left Behind, states were required to use criterion-referenced accountability systems. That is, after states established their benchmark for student proficiency, all schools were held accountable for ensuring all subgroups met those predefined targets. In contrast, the Every Student Succeeds Act requires states to set up normative accountability systems. Under normative systems, schools are compared to each other rather than to an absolute benchmark. States must identify their lowest-performing five percent of schools for comprehensive support, and schools with low-performing subgroups for targeted support, but the goal behind this type of accountability system is to triage and identify the schools in need of the most support.

Hypothetically, if COVID-19 affected all schools equally and test scores fell across the board, the state would end up identifying the exact same schools for support as it did in prior years. The main thing that matters is how any one school compares to all the others.

Of course, as explained above, COVID-19 is likely to lead to significant learning losses, and those losses will not hit all schools or communities equally. In that case, having a normative accountability system may help states prioritize supports and interventions that need it the most, but it won’t necessarily force states to identify a higher percentage of schools in need of support. If anything, state leaders should be willing to adjust their thresholds to identify a higher percentage of schools for additional funding and support.

There are also the questions of fairness and accuracy. With the spring 2020 waivers, states paused their accountability systems, and the schools that were identified for comprehensive and targeted support in 2019-20 will be required to continue their interventions into the 2020-21 school year. Without collecting new data and running their school support lists again in the summer of 2021, states will find themselves making consequential decisions while relying on increasingly outdated information.
What purposes are state tests meant to serve, and how can states continue to meet some or all of these in a post-COVID landscape?

In a normal year, states’ tests are intended to serve multiple overlapping purposes. They have instructional purposes, such as informing parents about their child’s performance, helping teachers diagnose the instructional needs of their students, and allowing school and district leaders to target resources and programmatic supports. State testing systems are used for accountability purposes by state and local policymakers to identify schools for recognition or support. Tests also provide the public with transparent information about educational outcomes. Finally, tests allow researchers the ability to evaluate programs and interventions. With the cancellation of the spring 2020 assessments, states will need to plan for how they might meet each of these purposes, likely through a combination of tests and other indicators.

How will states be able to inform parents and educators about student progress?

States typically administer one annual summative assessment in the spring, and those results are packaged and shared with parents and educators. One intended purpose for those results is to inform parents to enable them to make decisions about any outside supports their child might need, and to provide parents with a common language with which to communicate with their child’s teachers. Teachers and school administrators can also use the data to adjust their plans, ranging from lesson plans and curriculum decisions to staffing assignments. Without the 2020 spring testing data to play that role, states will need to find other ways to fill that void. For example, Tennessee is creating three optional assessment tools available for districts: a “start of the year checkpoint,” an online formative assessment, or a full-length interim assessment.14

For states that expect districts to fill that void on their own, will the state require or recommend that school districts administer their own diagnostic assessments? If so, what parameters or guidelines will states put on those district decisions? Will they curate an approved list of vendors, or leave it up to districts to decide whether they need them and, if so, which one to select?

Some states are already thinking creatively about how to provide information about student academic skills and gaps to parents, teachers, and educators. The Texas Education Agency, for example, worked with its testing contractor to create a set of “Optional Year-End Assessments” based on prior years’ test items. Families could request either an online login or a pencil-and-paper version mailed to their address. The state did not publish any results by district or school, but parents who chose the online option received their child’s results immediately. Texas made the test available on a voluntary basis to provide parents and district leaders with valuable information.
How will states ensure they test all students?

A provision in federal law requires states to test at least 95 percent of their students, and every school that fails to meet that threshold is flagged for additional support. The 95 percent participation rate provision is intended as a way to ensure that schools reach all students. It was waived in 2020 along with all the other testing provisions, but the lack of student connection with distance learning efforts this spring suggest that the 95 percent participation rate provision is only more important in a remote learning context. An unacceptably large number of students were disengaged with their schoolwork, and those students tended to be low-income, disadvantaged students without a computer and steady internet access.

Given those disparities, states should plan to administer their summative tests in the spring of 2021. Even if those tests must be administered in a remote environment to some or all students, states should be planning now for the accommodations or other adaptations to make that a reality. Can the state offer online or pencil-and-paper versions depending on the student’s educational setting at the time of testing? What remote proctoring or scheduling changes would the state need to consider?

These questions are answerable, if states are willing. In 2020, for example, while most states were canceling their tests for K-12 students, the College Board quickly created a scaled-down version of its Advanced Placement tests that millions of students took from their homes. While thousands of students reported challenges with submitting their tests online, the College Board claimed a 99 percent success rate overall. Similarly, the testing company behind the most popular teacher licensure test also created a version for candidates to take from home. Both of these cases involved older, self-motivated students, and states would need to offer additional supports when working with younger students, English learners, and students with disabilities, but these use cases at least illustrate a path forward is possible.

How will the state ensure it is meeting its obligations to students with disabilities and English learners?

Beginning in 1997 with the reauthorization of the Individuals with Disabilities Education Act (IDEA), Congress required states to include students with disabilities in state and district assessments and to report their participation rates and results. Under NCLB, Congress also required students with disabilities to count as a separate subgroup and held all schools accountable for their performance. In addition, students with disabilities often have specific goals in their Individualized Education Plans (IEPs) that are tied to their performance on state assessments. All students are required to receive instruction based on grade-level content standards, and state tests — or alternate assessments administered to no more than 1 percent of students with the most significant cognitive disabilities — help determine whether or not those goals are being met.
When COVID-19 forced schools to close, the 6 million students with disabilities were some of the most difficult student populations to serve well under distance learning. Under federal law, schools are legally required to provide a “free and appropriate public education” to all students with disabilities. While technology makes some services possible, other services can only be delivered in person. News reports at the time found many districts that were struggling to provide educational services that were appropriate for students with disabilities. To address these issues going forward, state leaders should be clearer about the expectations for schools to serve students with disabilities even under distance learning. Accountability for attendance and state test results can help uncover places where students are not being served equitably.

States are also required by federal law to administer an English language proficiency (ELP) assessment to their English learners. Those requirements were waived in 2020. Meanwhile, English learners faced some of the biggest disadvantages under distance learning given that they are more likely to struggle to access English-language content. English learners also come from households with less access to computers and the internet, adding another obstacle to their progress. In order to support English learners going forward, states will need objective information to identify the schools and districts that need the most help in reaching those students.

Can states still measure student growth?

Measuring how much students learn over time is one of the most equitable measures of student performance available. For example, growth data reveal trends and patterns that status measures of student proficiency rates might miss. Student growth measures also provide the best estimate of school and district contributions to student learning. In short, student growth information will be critical to understand the extent of the impact of school closures.

Under normal conditions, growth measures typically track how much students learn over the course of a single year, from one spring to the next spring. But with states canceling their 2020 spring assessments, there are valid questions about whether states can and will continue to measure student growth.

Fortunately, it is possible to measure growth over two years’ time. While the process isn’t the same as the more typical annual growth measures, states can produce valid, usable estimates of student growth by using the 2019 test as the baseline for assessments administered in the spring of 2021. This approach can be used with all types of growth models (e.g., value-add, student growth percentiles, etc.). This might seem unprecedented, but states routinely use two-year growth measures when prior year assessment data is unavailable or unusable. For instance, for years Massachusetts has calculated a growth measure based on a student’s eighth-grade test scores versus their tenth-grade test scores to calculate how much the student grew over ninth and tenth grade combined.
Other states have dealt with a very similar issue due to testing disruptions. Tennessee, for example, had technical issues with the rollout of its new online exam in 2016. The following year, the Tennessee Department of Education released guidance that data from 2015 would serve as the baseline for calculating value-added results in the spring of 2017. Research found that Tennessee’s two-year growth estimates were highly correlated to its standard one-year growth estimates.20

Two-year growth measures are not the norm, but they are a viable alternative to canceling 2021 growth measures.

**How will states measure student attendance?**

Thirty-six states and the District of Columbia hold schools accountable for chronic absenteeism.21 Although state definitions of “chronic absenteeism” vary somewhat, most states define a student as chronically absent if they miss 10 percent or more of school days. Prior to the coronavirus, states adopted policies to ensure consistency around what counted as an “absence,” and it was relatively easy to calculate whether a given student missed more than 10 percent of school days or not.

But it’s not clear what "attendance" means in a virtual environment. In 2020, most states waived their policies around minimum days of schooling, with some providing guidance to school leaders and others merely offering recommendations. With districts planning for various configurations of their school days to keep students socially distant, many district leaders are already contemplating having students physically in schools only a portion of the school week, while the rest of their time is spent in distance learning. Will states adopt guidelines for how much time students should be spending in distance learning for it to count toward a day of school? Or are states prepared to shift away from the seat-time measurement of student progress and allow schools to experiment with a competency-based approach? While such a radical departure may be welcomed for other reasons, state leaders would need to be prepared to adjust their assessment and accountability systems to allow for such experimentation, not to mention granting potential waivers for the experimentation itself.

**What other measures or rules will states need to adapt?**

The Texas example above also shows how states should think through the potential purposes and uses of testing results under a variety of circumstances. A state might use assessment results to target supports and interventions to students who need it, while also preserving a strong stance that results should do no harm to the participating students or educators. For example, states with laws requiring schools to hold back third-graders with low reading scores should be cognizant that the COVID-related school closures will likely mean more students fall into those categories in the coming years. States need a plan to deal with that, ideally one that addresses the underlying gaps rather than merely penalizing a higher percentage of students.
All states are required to measure and hold schools accountable for high school graduation rates. In 2020, many states waived or adjusted their graduation requirements to allow students to earn a diploma even if they had not passed high school exit exams or other requirements that would have required in-person engagement, such as work-based learning opportunities or internships. States with these policies on the books will need to consider whether those changes are extended for another year or if they need to make other adjustments that are more permanent.

Additionally, many states have built in other features of their accountability systems that may need to change. For example, states that normally average multiple years of results into one overall rating may want to exclude or reduce the weighting for 2020 results.

In the present moment, even a return to “normal” may seem daunting for our education systems. With COVID-19 death counts still rising and millions of Americans newly unemployed, the immediate economic consequences of the coronavirus are real and alarming enough. It’s understandable for teachers and school leaders busy dealing with the logistics of school operations to resent any attempts at accountability in the current environment.

But we respectfully disagree. The coronavirus has laid bare the inequities rife in our society, which our youngest members of society experience through reduced educational opportunities. Smart thinking from state leaders now could provide valuable feedback about which students and schools need the most help and quantify how much students are learning — or not — during our collective distance learning experiment. Smart decisions now could even serve as building blocks for more flexible, adaptive accountability systems in the future. Another year of missing data, lost opportunities, or a blanket waiver on accountability systems would be a disaster for our youngest, most vulnerable children.

| Smart thinking from state leaders now could provide valuable feedback about which students and schools need the most help and quantify how much students are learning — or not — during our collective distance learning experiment. |
Endnotes


18 While this section is mainly about test score growth, some states measure change over time on other indicators, such as chronic absenteeism. Those states could also use 2019 as their baseline and measure improvements from 2019 to 2021.


Acknowledgments

We would like to thank the Michael & Susan Dell Foundation for its financial support for this project, to Super Copy Editors, and to Five Line Creative for graphic design.

Any errors in fact or analysis are the responsibility of the authors alone.

About the Authors

Alex Spurrier
Alex Spurrier is a senior analyst at Bellwether Education Partners. He can be reached at alex.spurrier@bellwethereducation.org.

Chad Aldeman
Chad Aldeman is a senior associate partner at Bellwether Education Partners. He can be reached at chad.aldeman@bellwethereducation.org.

Jennifer O’Neal Schiess
Jennifer O’Neal Schiess is a partner at Bellwether Education Partners. She can be reached at jennifer.schiess@bellwethereducation.org.

Andrew J. Rotherham
Andrew J. Rotherham is co-founder and a partner at Bellwether Education Partners. He can be reached at andy.rotherham@bellwethereducation.org.

About Bellwether Education Partners

Bellwether Education Partners is a national nonprofit focused on dramatically changing education and life outcomes for underserved children. We do this by helping education organizations accelerate their impact and by working to improve policy and practice. Bellwether envisions a world in which race, ethnicity, and income no longer predict opportunities for students, and the American education system affords all individuals the ability to determine their own path and lead a productive and fulfilling life.