RAND CORPORATION

Research Report

JONATHAN SCHWEIG, MEGAN KUHFELD, MELISSA KAY DILIBERTI, ANDREW McEACHIN, LOUIS T. MARIANO

Changes in School Composition During the COVID-19 Pandemic

Implications for School-Average Interim Test Score Use

KEY FINDINGS

- Participation in Measures of Academic Progress (MAP) Growth assessments was lower in 2020–2021 than in pre-pandemic years. Of students taking MAP assessments in spring 2019, 42 percent were assessed again in spring 2021, and 21 percent of these test-takers attended the same schools across these two school years.
- Within and among districts, there was wide variability in the percentage of students who attended the same schools and participated in MAP Growth assessments over two academic years.
- Participation in MAP Growth assessments was uneven in 2020–2021. In particular, students of color were less likely to have attended the same schools and participated in MAP Growth assessments over two academic years than were White students.
- Historically higher-achieving students who participated in assessments in a given year were generally more likely than their peers to have attended the same schools and participated in MAP Growth assessments over two academic years.
- Schools serving high-poverty communities and communities vulnerable to coronavirus disease 2019 (COVID-19) had systematically fewer students attend the same school and participate in MAP Growth assessments over two academic years than other schools.

The novel coronavirus disease 2019 (COVID-19) pandemic has had a profound impact on the lives of young people. The pandemic disrupted almost every aspect of kindergarten through 12th-grade education in the United States and exacerbated long-standing and profound racial and socioeconomic disparities in access and opportunity in U.S. public schools (Kaufman and Diliberti, 2021; U.S. Department of Education, Office for Civil Rights, 2021). Schools closed to in-person instruction in March 2020 (Peele and Riser-Kositsky, 2021), and many students were still receiving at least some remote instruction well into the 2020–2021 school year (Institute of Education Sciences, undated).[1] As a result, nearly 55 million school-aged children in the United States had reduced access to the social and academic supports that are typically

Abbreviations

COVID-19	coronavirus disease 2019
FRPL	free or reduced-price lunch
MAP	Measures of Academic Progress
PVI	Pandemic Vulnerability Index
U.S. ED	U.S. Department of Education

provided by schools (García and Weiss, 2020). Students were isolated from their friends and teachers (Digital Promise, 2021), and parents and teachers reported that the pandemic made it more difficult to participate in sports, music, arts, and other after-school activities (Digital Promise, 2021).

Although the spread of the highly transmissible Delta variant throughout summer 2021 raised many concerns about safely starting the 2021–2022 school year, officials largely held fast to the notion that a full return to in-person learning was a high priority. Early data suggest that most schools are back to full-time in-person learning for the 2021–2022 school year (Herold, 2021). Accordingly, federal and state policy has pivoted to focus on *restart* and *recovery*; school systems are implementing policies and programs aimed to keep students safe from COVID-19 and to help promote students' learning and social and emotional well-being (Centers for Disease Control and Prevention, 2021; Education Policy Innovation Collaborative, 2021).

Many educators and policymakers advocated for statewide summative assessment programs to be restarted in the 2020–2021 school year to evaluate the state of teaching and learning and to inform policies and practices for COVID-19 recovery (Bruno and Goldhaber, 2021; Caprariello, 2021; Council of Chief State School Officers, 2020). This position was incorporated into federal COVID-19 response policies. Specifically, unlike in the 2019–2020 school year, when the U.S. Department of Education (U.S. ED) offered blanket assessment waivers to all 50 states, in the 2020–2021 school year, U.S. ED took the position that statewide summative tests would play an important role in assessing school and student progress, identifying schools in need of support, and informing resource allocation decisions (U.S. ED, 2021b). However, U.S. ED also emphasized the importance of flexibility in spring 2021 assessment, creating an opportunity for officials to prioritize the use of benchmark or interim assessments to guide decisionmaking (Bruno and Goldhaber, 2021; U.S. ED, 2021a).

In addition to individual student test scores, school-aggregate scores from interim, benchmark, and state summative tests will almost certainly play a key role in determining which schools have responded successfully to COVID-19–induced disruptions and, likewise, in identifying schools in which students may have been disproportionately affected. School-aggregate test scores are routinely used by state and local education agencies to inform resource allocation decisions and to plan staff hiring. Through the Coronavirus Aid, Relief, and Economic Security Act, the Coronavirus Response and Relief Supplemental Appropriations Act, and the American Rescue Plan Act (Public Law 116-136, 2020; Public Law 116-260, 2020; Public Law 117-2, 2021), U.S. ED has allocated funding to state and local school districts to implement programs that address disrupted learning, and school-aggregate test scores are likely to play an important role in evaluating the effectiveness of these programs (Education Analytics, 2021).

However, using school-aggregate test scores to determine which schools or districts are handling COVID-19 restart and recovery well and which need additional support relies on the assumption that differences in aggregate test scores can be accurately interpreted as representing real and meaningful differences in school progress and performance. There are serious concerns about the accuracy of such interpretations even under routine schooling conditions (e.g., Luyten and de Wolf, 2011; Polikoff, 2019), but the COVID-19 pandemic may exacerbate these issues and further compromise the comparability of aggregate-level test scores. Importantly, differences in aggregate test scores likely reflect not only true pandemic impacts on student learning but also other factors that have nothing to do with school progress or performance. These include fundamental questions about differences in assessment mode (e.g., remote versus in-person assessment), differences in instructional mode, differences in opportunity to learn, and the differential emotional and psychological effects of COVID-19–related stressors.

In this report, we investigate one specific issue that has a contaminating influence on school-aggregate test scores: changes in test-taking populations and school composition over time. There was an increased propensity during the pandemic for families to move residences, and such residential changes are often accompanied by changes in school or school district enrollment (Bouzaghrane et al., 2021; Jacobson, 2021b; Mahnken, 2021). Furthermore, early evidence suggests that, in 2021, there were steep declines in the number of students who participated in spring state summative tests. Although some states reported overall participation rates above 90 percent, other states saw overall participation rates as low as 10 percent (see Table 1).

In our study, using nationwide analyses of NWEA's Measures of Academic Progress (MAP) Growth assessments—online, computer-adaptive tests designed to measure academic growth for students in kindergarten through 12th grade—we sought to address the following questions:

- What percentage of students taking MAP Growth assessments in spring 2019 were assessed again in spring 2021, and what percentage of these students attended the same schools across these two school years? How does this compare with pre-pandemic patterns?
- Is there variability in the percentage of students who attended the same schools in these two school years? To what extent does this percentage vary within and among districts?
- Among students who took MAP Growth in spring 2019, to what extent is spring 2021 test participation associated with the student, school, and community characteristics of gender, race/ethnicity, poverty, and pandemic vulnerability?

To address these questions, we constructed two-year match rates (Ho, 2021) for pre-pandemic and pandemic periods to estimate the percentage of students who attended the same schools and participated in MAP Growth assessments over two academic years. We then used these match rates to explore how the school composition of MAP Growth test-takers has shifted over time. The objective of this study is to illustrate the extent to which lower or differential participation in spring 2021 testing as compared with prior years can compromise the comparability of aggregate-level test scores over time or among schools in the context of the COVID-19 pandemic. By detailing systematic shifts in school composition and test participation during the pandemic, this study potentially informs stakeholders about the extent to which school-level test score aggregates based on benchmark assessments like NWEA's MAP Growth provide accurate information about school progress and performance during the 2020–2021 school year. Detailing these shifts provides empirical evidence that highlights important contextual considerations that might support the use of school-aggregate test scores to support school recovery. Although the reliance of this study on MAP Growth data may limit the generalizability of our conclusions to statewide summative testing contexts, the findings in this report contribute to broader national conversations about whether and when it is appropriate to use school-aggregate test scores to appraise and monitor school progress and performance throughout the pandemic. Our findings highlight the importance of

The findings in this report contribute to broader national conversations about whether and when it is appropriate to use school-aggregate test scores to appraise and monitor school progress and performance throughout the pandemic.

TABLE 1

State Summative Tests and School-Level Score Usage: 2020–2021

State	Participation Rate (percentage)	School-Level Score Reports	Information About School-Level Score Usage	Specific Uses
AK	64.0	Yes	No	
AL	93.0	Yes	No	Plan professional learning and evaluate programs
AZ	85.0	Yes	Yes	Target support and resources
CO	51.0–72.0	Yes	Yes	Target support and monitor school recovery
CT	93.5	No	No	
DE	60.0	Yes	Yes	Address short- and long-term learning needs
FL	93.0	Yes	No	
GA	55.0–79.0	Yes	Yes	Measure knowledge and skill mastery
IA	93.0	Yes	No	
ID	N/A	Yes	Yes	Inform the public and identify growth opportunities
IN	96.5	Yes	Yes	Inform the public
KY	76.0–89.0	Yes	Yes	Address pandemic-induced learning disruptions
LA	97.5	Yes	Yes	Guide policy and resource allocation decisions
MA	97.0	Yes	Yes	Assess student learning and improve schooling
MI	< 75.0	Yes	Yes	Target support and resources
MN	78.2	Yes	Yes	Support local decisionmaking
MO	90.5	No	Yes	Identify schools that need additional support
MS	96.9	Yes	Yes	Measure knowledge and skill mastery
MT	91.0	No	Yes	Identify opportunity differences among schools
NC	93.0	Yes	Yes	Inform the public and guide instruction
NM	10.0	No	No	
OH	N/A	No	Yes	Inform the public
OK	92.0	No	Yes	Monitor academic growth and performance
OR	30.0	No	No	
SC	87.9	Yes	No	
TN	80.0–95.0	Yes	Yes	Identify excelling schools and those in need of additional support; determine priority exit status
TX	87.0	Yes	Yes	Identify schools that need additional support
VA	76.5	Yes	Yes	Address pandemic-induced learning disruptions
WV	91.0	Yes	Yes	Analyze individual student performance
WY	96.0	Yes	No	

SOURCES: Participation rates are taken from Center on Reinventing Public Education, 2021. School-level score reports, information about school-level score usage, and specific uses are from our analysis of each state's website; see, for example, Colorado Department of Education, 2021; Michigan Department of Education, 2021; and West Virginia Department of Education, 2021.
NOTE: N/A = not applicable.

transparently reporting information on test participation to "advance the goal of accurate score interpretations and fair trend comparisons among schools and districts" (Ho, 2021, p. 1).

We proceed with four main sections. First, we provide background information on the assessment context. We detail the problems caused by changes in test-taking populations, particularly for school-level analyses, and provide examples of state policies for score reporting. Second, we detail our research questions and describe our sample and methods. Third, we present the results of our investigations. We conclude with a discussion of the implications of our findings for decisionmaking during the COVID-19 crisis.

Study Background and Context

In response to the COVID-19 pandemic, all 50 states closed schools to in-person instruction in spring 2020 (Peele and Riser-Kositsky, 2020). By April 2020, U.S. ED had granted blanket assessment and accountability waivers to all 50 states to allow state and local education agencies to bypass annual summative state testing requirements under the Every Student Succeeds Act—likely with the assumption that students would be back to school in person for the 2020–2021 school year (Gewertz, 2020). Accordingly, no states conducted summative statewide testing in spring 2020. Additionally, most school districts decided to skip interim or benchmark assessments that normally would be administered in the spring (for example, less than 5 percent of students who are normally assessed in the spring via MAP Growth were administered tests in spring 2020) (Wise, Kuhfeld, and Cronin, 2021).

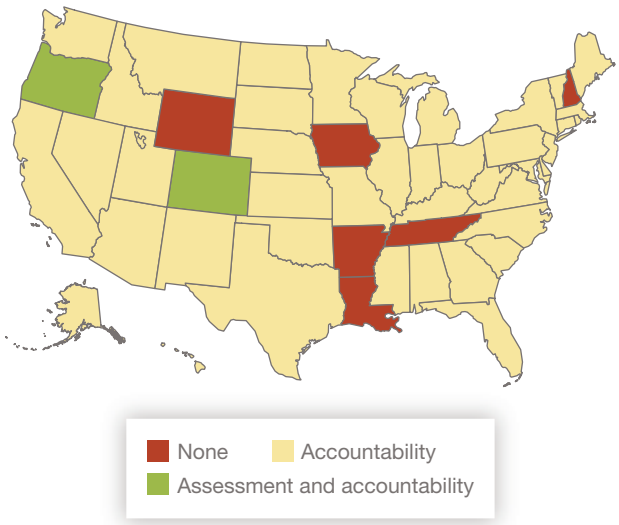
However, by February of the following school year (2020–2021), most U.S. public school students were still learning remotely, and students of color were significantly more likely than White students to still be receiving at least some remote instruction (National Center for Education Statistics, 2021). In this public health and policy context, U.S. ED announced an updated assessment and accountability waiver policy. Unlike during the 2019–2020 school year, U.S. ED did not give blanket permission for states to cancel state summative tests in 2020–2021. Instead, U.S. ED took the position that "state assessments and accountability systems play an important role in advancing educational equity, identifying student needs, and targeting the resources to address them" (U.S. ED, 2021a).

Although U.S. ED required states to conduct state summative testing in some form, it also offered more flexibility than usual. For example, U.S. ED offered accountability waivers that, among other provisions, removed the requirement that states use 2020–2021 data to identify schools for support and instead required states to resume school identification based on 2021–2022 data. U.S. ED also offered states flexibility in terms of the timing, administration mode, and length of state summative assessments and the proportion of students to be tested (U.S. ED, 2021a).

Ultimately, 45 states submitted accountability waiver requests, and all these requests were approved (see Figure 1) (U.S. ED, Office of Elementary and Secondary Education, 2021). Twelve states and the District of Columbia requested waivers of summative state assessments, although only three such waivers were ultimately approved. In states that did not have

FIGURE 1

Federal Assessment and Accountability Waivers, 2020–2021 School Year



SOURCE: U.S. ED, Office of Elementary and Secondary Education, 2021.

NOTE: The District of Columbia was also granted an assessment and accountability waiver but is too small to be displayed in this figure.

assessment waivers, there was considerable variability in approaches to testing and the amount of emphasis that was placed on ensuring participation in statewide summative assessments. The result of these policies is that student assessment data were widely collected in spring 2021 but under conditions that potentially limit the comparability of these data over time or among schools.

Many uses of test score data (whether from interim, benchmark, or summative assessments) focus on individual students' learning. For example, school leaders use test data to evaluate whether individual students are ready for grade advancement or graduation (e.g., Dougherty, 2015; Hamilton et al., 2009; McEachin, Domina, and Penner, 2020) or to inform students' course placement recommendations and to select students for specialized programming (e.g., Beaver and Weinbaum, 2015; Means, Padilla, and Gallagher, 2010). However, school-aggregate test score data are also an essential source of information. Under regular (pre-pandemic) schooling conditions, school-aggregate scores from state summative assessments play a central role in state accountability systems and are routinely reported on state websites, where they can be used by officials to monitor population trends and patterns and to compare and monitor school or subgroup performance (Keng and Marion, 2020). School-aggregate scores from a variety of assessments, including interim, benchmark, and state summative tests, are also routinely used by state and local education agencies to inform resource allocation decisions, to plan staff hiring, and to evaluate interventions.

Student assessment data were widely collected in spring 2021 but under conditions that potentially limit the comparability of these data over time or among schools.

Although states' formal accountability systems were suspended for the 2020–2021 school year, school-aggregate scores can still play an important role in monitoring school performance because school-level assessment information is needed to address COVID-19–related unfinished learning and opportunity gaps (Hamilton and Ercikan, 2022). U.S. ED recommended that system officials use aggregated scores from state summative assessments to support restart and recovery, and this recommendation has been taken up by many school systems. Other states have focused instead on using school-aggregate scores from benchmark or interim assessments like MAP Growth to support restart and recovery. California implemented a statewide policy that allowed districts to use the "best assessment tool available for the local context" (California Department of Education, 2021), and several large districts, including the Los Angeles Unified School District, de-emphasized participation in statewide summative assessment and instead relied on interim and benchmark assessments like MAP Growth to inform resource allocation decisions and guide restart and recovery strategies (Stokes, 2021). Michigan's "Return to Learn" legislation emphasized the use of benchmark assessments like MAP Growth to monitor and promote restart and recovery (Education Policy Innovation Collaborative, 2021). New Jersey delayed spring 2021 assessments to fall 2021 and administered interim assessments that were specially designed to diagnose student academic needs and plan for resource allocation (New Jersey Department of Education, Office of Assessments, 2021).

State recommendations surrounding school-level test score use in the context of COVID-19 fall into three main categories. First, several states have noted that being transparent in their reporting of school-level data from state summative tests is important as a matter of civic responsibility and informing the general public about the current state of schools and schooling. Accordingly, of the 30 states that have released summative test score data as of the writing of this report, 23 have made school-aggregate test

score data publicly available, and 22 have provided explicit guidance on the intended uses of these school-level test scores (Table 1).[2] Second, states have noted that school-level test scores from benchmark, interim, and summative assessments are important for state and local district officials to identify schools where students have responded successfully to COVID-19–induced disruptions and, likewise, for officials to identify schools where students might need additional support. States have also noted that school-level test scores are helpful for identifying where opportunity gaps are persistent or have been exacerbated by COVID-19. Third, states have noted that school-level test data from benchmark, interim, and summative assessments are important for informing resource allocation decisions, including the allocation of Elementary and Secondary School Emergency Relief funds (federal funds intended to help schools address COVID-19–related challenges and recovery) (e.g., Keng and Marion, 2020; U.S. ED, 2021b). For example, school-level data are expected to play a key role in informing staffing decisions and in evaluating programs and interventions designed to mitigate the academic impacts of COVID-19 (Jacobson, 2021a).[3]

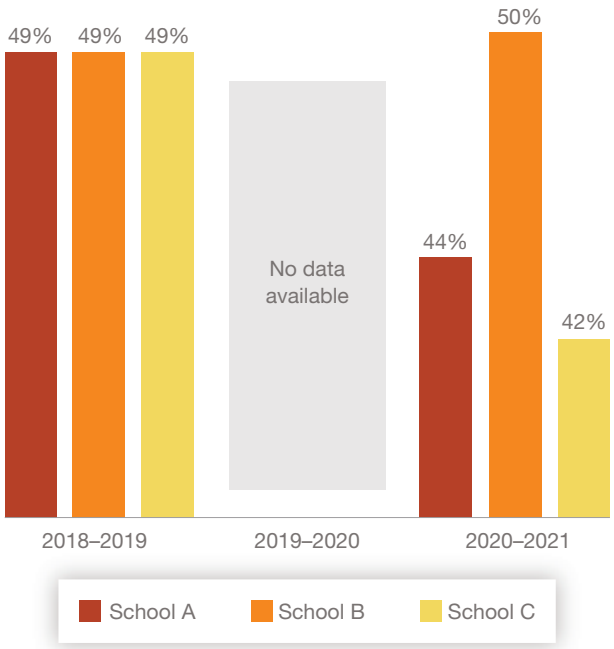
Accomplishing these goals using school-aggregate test scores requires measurement processes that allow fair and accurate comparisons among schools and over time (Ho, 2021). However, low or uneven participation in spring 2021 testing can result in changes to school composition that compromise the accuracy of such comparisons (Keng and Marion, 2020).

Changes to School Composition

Comparing school-average test scores (including proficiency percentage) or interpreting a change in such scores over time requires assumptions about the compositional stability of the test-taking populations at each school, even if a school's overall composition is unchanged. To illustrate this, consider the proficiency percentages for three hypothetical schools (School A, School B, and School C) that each enrolled a constant set of 100 students from 2018–2019 (before the pandemic) to 2020–2021 (during the pandemic). In Figure 2, School A appears to have a performance

FIGURE 2

Changes in Proficiency Rates for Three Hypothetical Schools



decline from 2018–2019 to 2020–2021 of about 5 percentage points. School B appears to have a slight performance increase (up 1 percent). School C appears to have had the largest decline of the three schools (about 7 percent). However, Figure 2 conceals three important pieces of information. In School A, all 100 students participated in assessments in 2020–2021. In School B, the eight lowest-scoring students in 2018–2019 did not participate in assessment in 2020–2021. In School C, the eight highest-scoring students did not participate in assessment in 2020–2021. How do we interpret the progress of these schools over time or their performance relative to other schools in the state? This simple exercise illustrates how low and uneven test participation creates the possibility that observed differences in school-level test scores among schools or over time might reflect decreases in the participation of traditionally lower-performing student groups, students newly opting out, real changes in student achievement, or a mixture of the three.

The issue of stability in test-taking populations is important even in school years not affected by COVID-19 (e.g., Polikoff, 2019). Research on school mobility, for example, suggests that significant num-

bers of students change schools or districts in a given year (Welsh, 2017), and school mobility is particularly common in districts with high incidences of poverty and large proportions of students of color (Lleras and McKillip, 2017). Students may also opt out of testing, and some studies show that White students and students from higher-income districts are more likely to opt out of testing, although other studies show that opt-outs have become increasingly popular in communities of color (Bryant, 2016; Schweig, 2016). Literature on student opt-outs suggests that even small compositional changes can have dramatic effects on school-aggregate test scores. One study showed that removing 10 percent of low-achieving students from the calculation of a school-aggregate test score could cause a school that would be identified as in need of support to be identified as making satisfactory progress (Beaver, Westmaas, and Sludden, 2014; Cremata, 2019).

This issue is likely to be particularly salient in making inferences about COVID-19 impacts using school-aggregate test data. Because federal accountability waivers removed the 95-percent student participation mandate for state summative testing in the 2020–2021 school year, it is likely that there will be more-substantial numbers of students who opted out of state summative testing. Participation rates varied widely within and among the 22 states in Table 1 that have released spring 2021 state summative test scores. Some states reported overall participation rates above 90 percent, and other states saw overall participation rates as low as 10 percent. In some states, such as Colorado and Georgia, participation varied substantially among districts.

There is evidence of systematic differences in state summative test participation during the pandemic based on student characteristics. In Ohio, participation declines for Black students were nearly three times larger than for their White peers (Barnum, 2021). In Michigan, Black students and economically disadvantaged students were less likely to participate in assessment than their White and more economically advantaged peers (Education Policy Innovation Collaborative, 2021).

Similar patterns have been documented for interim and benchmark assessments. In Michigan, participation rates in such assessments were around 75 percent (Education Policy Innovation Collaborative, 2021). An analysis using MAP Growth data (Johnson and Kuhfeld, 2020) found systematic demographic differences across subjects and grades in test participation prior to and during the pandemic. Furthermore, a larger fraction of students who did not participate were students of color, students with lower fall 2019 achievement, and students in schools serving a larger proportion of economically disadvantaged students.

Additionally, there might be substantial and atypical numbers of students who are not in school who otherwise would be enrolled, above and beyond typical student mobility rates. Absenteeism rates in some urban districts doubled in 2020–2021 (Kurtz, 2020; O'Donnell, 2020). Public school enrollment for kindergarten through 12th grade declined by nearly 3 percent across the United States compared with 2019–2020 (National Center for Educational Statistics, 2021). In some communities, economic hardships induced by COVID-19 forced families to change school districts. In other (often more economically advantaged) communities, parents who were dissatisfied with remote learning in public school settings enrolled in charter schools or opted out of public education entirely (Bouzaghrane et al., 2021; Jacobson, 2021b; Mahnken, 2021).

Given the context of COVID-19, making accurate judgments about school recovery depends on a thorough understanding of the extent to which changes in aggregate scores reflect meaningful performance changes, as opposed to low or uneven student test participation.

Study Approach

For this study, we used individual student–level MAP Growth data from NWEA's anonymized longitudinal Growth Research Database to construct two-year match rates (Ho, 2021) for pre-pandemic and pandemic periods. These data were used to characterize changes in student test participation. We explored the implications for the use of school-level test score aggregates in identifying schools where students have responded successfully to COVID-19–induced disruptions, identifying schools where students might need additional support, and informing resource

allocation decisions. MAP Growth assessments differ from state summative tests in several ways, including standards alignment, test content, administration, and score interpretation, and we caution that these differences may limit the extent to which our analyses can be generalized to statewide summative testing contexts. However, as we detailed in the previous section, given that some states were granted assessment waivers for 2020–2021 and other states (including California) that were not formally granted waivers did not offer statewide assessments, benchmark and interim assessments like MAP Growth will play a role in state and local plans to track and monitor restart and recovery. This makes the MAP Growth assessments suitable for our research questions. Additional limitations of these data are discussed later in this section.

Study Sample

All analyses included in this report used student-level data from spring MAP Growth reading assessments.[4] We defined a set of inclusion criteria to determine which student records would be retained in a common analytic file used for all analyses. Although some analyses were conducted at the student level, and others were conducted at the school level, the population of students included in the analytic sample remained the same. Our analytic sample consisted of data from three school years: 2016–2017, 2018–2019, and 2020–2021. Because of the absence of spring 2020 MAP Growth reading data, we examined two two-year spans: spring 2017 to spring 2019 (which we refer to as the *pre-pandemic period* for the purpose of comparison), and spring 2019 to spring 2021 (which we refer to as the *pandemic period* for examining the impacts of COVID-19 on test participation).

From these assessment data, we created two analytic files: one file that linked schools from spring 2017 to spring 2019 and one file that linked schools from spring 2019 to spring 2021. In addition to MAP Growth scores, NWEA collects information about student grade level, race, ethnicity, and gender. To be included in our analytic files, a school needed to have at least ten students with a test score at baseline and at least one student with a test score at follow-up.

Benchmark and interim assessments like MAP Growth will play a role in state and local plans to track and monitor restart and recovery.

Approximately 18 percent of the students were ineligible for the pre-pandemic analytic file, and approximately 36 percent of the students were ineligible for the pandemic analytic file. This increase in ineligibility was driven by schools that opted out of administering MAP Growth entirely in spring 2021.

Table 2 provides descriptive statistics by grade for the pre-pandemic and pandemic analytic samples after these inclusion restrictions were implemented. Our analytic sample linking schools from spring 2017 to spring 2019 included 3,213,068 third- to eighth-grade students in 10,924 public schools (about 15 percent of the approximately 77,000 U.S. public schools serving this grade range). Our analytic sample linking schools from spring 2019 to spring 2021 included 2,484,142 third- to eighth-grade students in 8,511 public schools. Overall, the pandemic sample was 51-percent male, 52-percent White, 13-percent Black, 4-percent Asian, and 17-percent Hispanic, and similar demographic patterns were observed across grade levels and in the pre-pandemic year. Even after we placed the restrictions on our analytic sample, it remained representative of the U.S. population of public school students. However, the NWEA sample had a slight overrepresentation of White students and a slight underrepresentation of Hispanic students compared with the national population.

Instruments and Measures

Longitudinal Match

Following Ho, 2021, we created a student-level match indicator, defined as y_{ij}, by determining whether a

TABLE 2

Descriptive Statistics for the Study Sample

Demographic Characteristic	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
Analytic sample (2017)							
White	0.47	0.48	0.50	0.51	0.51	0.52	0.50
Black	0.17	0.16	0.15	0.14	0.14	0.15	0.15
Hispanic	0.19	0.19	0.18	0.18	0.18	0.18	0.18
Asian	0.04	0.04	0.04	0.05	0.04	0.04	0.04
Other	0.13	0.12	0.13	0.12	0.12	0.12	0.12
Male	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Number of students	526,809	545,460	581,778	548,584	525,552	484,885	3,213,068
Analytic sample (2019)							
White	0.49	0.51	0.52	0.54	0.54	0.54	0.52
Black	0.14	0.14	0.13	0.13	0.13	0.13	0.13
Hispanic	0.19	0.18	0.17	0.17	0.17	0.17	0.17
Asian	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Other	0.14	0.13	0.13	0.13	0.13	0.12	0.13
Male	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Number of students	401,335	417,132	459,968	434,018	404,280	367,409	2,484,142
U.S. population of public school students (2019–2020)							
White	0.45	0.45	0.45	0.46	0.46	0.46	0.46
Black	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Hispanic	0.28	0.28	0.28	0.28	0.28	0.28	0.28
Asian	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Other	0.06	0.06	0.06	0.06	0.06	0.05	0.06
Male	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Number of students	3,709,324	3,728,665	3,824,491	3,918,899	3,940,313	3,884,836	23,006,528

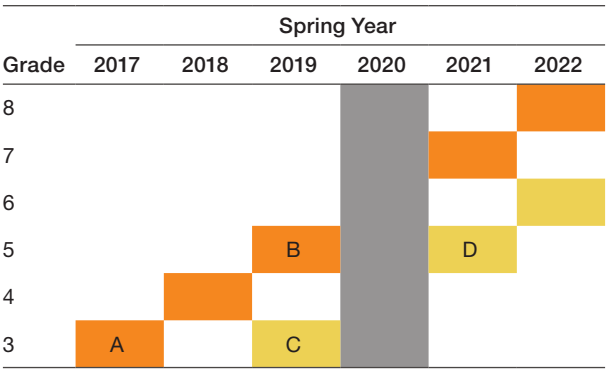
student *i* in school *j*, present in the baseline year (i.e., 2017 for the pre-pandemic period and 2019 for the pandemic period) in a specific school, made expected grade progression and remained enrolled in the same school in the follow-up year (i.e., two school years later). If both of these criteria were met, the student-level match indicator took on a value of 1. If any of these criteria were not met, the student-level match indicator took on a value of $y_{ij} = 0$. Note that there are three primary reasons that a match indicator would take on a value of $y_{ij} = 0$: (1) the student was not assessed in the follow-up year, (2) the student was assessed in the follow-up year but did not make expected grade progression, or (3) the student was assessed in the follow-up year but was enrolled at a different school. These criteria were applied to the data file linking schools from spring 2017 to spring

2019 and, separately, to the data file linking schools from spring 2019 to spring 2021.

Table 3 illustrates the logic of the match indicator for two cohorts of third-grade students who are enrolled at a single school. If a student appears as a member of set A *and* as a member of set B, that student is assigned a match value of $y_{ij} = 1$ in the pre-pandemic data file. Likewise, if a student is a member of set C and set D, that student is assigned a match value of $y_{ij} = 1$ in the pandemic data file.

TABLE 3

Longitudinal Match Rates for Two Periods for an Individual School



SOURCE: Adapted from Ho, 2021, p. 2.
NOTE: Orange cells represent the year-to-year progress of a single cohort matched in the pre-pandemic period; yellow cells represent the year-to-year progress of a cohort matched in the pandemic period; and gray cells represent the COVID-19–interrupted assessment year.

MAP Growth Assessments

The MAP Growth assessments are computer adaptive; are vertically scaled across grades; and measure student achievement in math, reading, language usage, and science using items aligned to state standards. This scaling allows educators to make various comparisons over time and among school districts. MAP Growth is used for various purposes within and among districts, including as a measure of academic growth and student goal-setting, as a component of curricular and programmatic placement decisions, as a component of admissions decisions for selective-enrollment high schools, as a universal screener for intervention programs, and as a component of teacher evaluation and school accountability systems. MAP Growth assessments are linked to most state accountability assessments (NWEA, undated), allowing educators to use MAP Growth results throughout the school year to predict a student's likelihood of being proficient on the end-of-year state accountability tests. MAP Growth assessments are administered at multiple points during the school year; students typically take them in the fall, winter, and spring of each school year.

School Context

NWEA does not collect information on school locale (urban, suburban, town, or rural) or economic disadvantage directly from participating students and schools. We obtained these data from the 2019–2020 Common Core of Data files (National Center for Education Statistics, undated). We used free or reduced-price lunch (FRPL) eligibility as a measure of school poverty. Because of the source of the data, we were unable to identify the FRPL eligibility of individual students and instead used school-level information in our analyses. *School-level FRPL eligibility* describes the percentage of students in each school who are eligible to receive free or subsidized meals as part of the National School Lunch Program (a federal program administered by the U.S. Department of Agriculture [USDA] that is designed to provide healthy meals to children from low-income families). As a note, some schools and districts use the Community Eligibility Provision and consider all of their students to be eligible for FRPL. Because we do not have an alternative measure of school poverty, all schools using the USDA provision have 100-percent FRPL eligibility in our sample. Overall, the sample closely aligns with the characteristics of U.S. public schools, including the percentage of FRPL eligibility and the percentage of urban, rural, and suburban schools (Table 4).

Vulnerability to COVID-19

We used a county-level measure of community vulnerability called the COVID-19 Pandemic Vulnerability Index (PVI). The PVI is published daily by the National Institute of Environmental Health Sciences and was designed to monitor disease trajectories and communicate local vulnerability (Marvel et al., 2021; National Institute of Environmental Health Sciences, undated b). The PVI is composed of four domains: infection rate, population concentration, interven-

TABLE 4

School-Level Descriptive Statistics for the Study Sample

Characteristic or Measure	Analytic Sample (2017)				Analytic Sample (2019)				U.S. Population of Public Schools			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
FRPL	0.52	0.28	0.00	1.00	0.52	0.28	0.00	1.00	0.55	0.28	0.00	1.00
PVI	0.50	0.06	0.23	0.62	0.50	0.06	0.25	0.62	0.51	0.06	0.23	0.64
Urban	0.28	0.45	0.00	1.00	0.23	0.42	0.00	1.00	0.28	0.45	0.00	1.00
Suburban	0.34	0.47	0.00	1.00	0.35	0.48	0.00	1.00	0.31	0.46	0.00	1.00
Town	0.11	0.32	0.00	1.00	0.12	0.32	0.00	1.00	0.13	0.34	0.00	1.00
Rural	0.27	0.44	0.00	1.00	0.31	0.46	0.00	1.00	0.28	0.45	0.00	1.00
MAP score	0.00	0.47	–3.22	1.66	–0.01	0.41	–2.97	1.31	—	—	—	—

SOURCES: Data on school locale and FRPL are from the Common Core of Data (2019–2020) (National Center for Education Statistics, undated). PVI is from National Institute of Environmental Health Sciences, undated.

NOTES: Number of schools = 10,924 (2017); 8,511 (2019). SD = standard deviation. — = statistics are not available for the U.S. population.

tion measures (e.g., social distancing, testing), and health and environment. The fourth domain—health and environment—contains information on the percentage of the population that identifies as Black or American Indian. (For more information about how this index is constructed, see National Institute of Environmental Health Sciences, undated a). We pulled the PVI each day from February 28, 2020 (the first day the PVI was published), to April 7, 2021 (our cutoff date for beginning analysis), and took the average. Overall, the PVI scores for the schools in our sample align closely with those of all U.S. public schools.

Analytic Methods

To describe the percentage of students taking assessments in spring 2019 who were assessed in spring 2021, we created longitudinal student-level matches using NWEA MAP Growth data from spring 2019 and spring 2021 (Ho, 2021). Using these data and the student match indicators described earlier in this section, we derived a longitudinal match rate (m_{21j}) for school j during the pandemic period as

$$m_{21j} = \frac{N_{CD}}{N_C}, \text{ where } N_{CD} = \sum_{i=1}^{N_C} y_{ij} = 1_{\{i \in C \cap D\}},$$

where m_{21j} is a match rate for school j, which is the ratio of the number of students present in the school in both the baseline and follow-up years (N_{CD}), calculated using the dichotomous student-level match indicator for student i in school j described above, and N_C is the number of tested students present in the school in the baseline year.[5] An overall match rate for J schools can thus be calculated as

$$M = \frac{1}{J} \sum_{j=1}^{J} m_{21j}.$$

For established schools with typical grade configurations (elementary schools spanning kindergarten through fifth grade and middle schools spanning sixth through eighth grade) and equal numbers of tested students enrolled in each grade, the expected maximum value of m_{21} would be 33 percent because many students would not be enrolled in the same school at follow-up because of standard promotional moves.

To contextualize and interpret M, we calculated a similar match rate for a two-year pre-pandemic span between spring 2017 and spring 2019:

$$m_{19j} = \frac{N_{AB}}{N_A}, \text{ where } N_{AB} = \sum_{i=1}^{N_A} y_{ij} = 1_{\{i \in A \cap B\}},$$

where m_{19j} is the match rate for school *j*. An overall match rate for *J* schools is likewise:

$$\tilde{M} = \frac{1}{J}\sum_{j=1}^{J} m_{19j}.$$

In addition to these longitudinal match rates, we calculated a measure of test participation across all schools, which was taken as the overall proportion of students with valid test scores in the baseline year who also had test scores in the follow-up year, regardless of their school affiliations.

It is important to note that both our longitudinal match rates and our participation rates take baseline-year test participants as the denominator for the purposes of calculation. This deviates from other cross-sectional test participation rates commonly reported on state websites, which report test participation as the fraction of enrolled students tested. However, we did not have 2020–2021 enrollment data available to construct participation rates in this way. It is also important to note that we could only tell whether a student remained in the same school if they had test data. Thus, a student that could not be matched to a school across time might have changed schools but might also have opted out of assessment in the follow-up year.

To investigate the extent to which pandemic period match rates varied within and among districts and states, we used a three-level regression model with schools contained within districts contained within states:

$$m_{jds} = \alpha + v_{00s} + u_{0ds} + e_{jds},$$

where m_{jds} is the match rate for school *j* in district *d* in state *s*. The overall mean match rate is captured by α, and v_{00s}, u_{0ds} and e_{jds} are state, district, and residual random effects, with mean zero and variances τ_β, τ_π, and σ^2, respectively. The proportion of total variance (Raudenbush and Bryk, 2002) in match rate among states can be expressed as

$$\frac{\tau_\beta}{\tau_\beta + \tau_\pi + \sigma^2},$$

and the proportion of total variance among districts (within states) can be expressed as

$$\frac{\tau_\pi}{\tau_\beta + \tau_\pi + \sigma^2}.$$

We used similarly defined models to investigate pre-pandemic match rates.

To examine the extent to which longitudinal matches differ systematically depending on the student, school, and community characteristics of race/ethnicity, poverty, pandemic vulnerability, and urbanicity, we used a multilevel linear probability model:

$$y_{ijds} = \alpha + \beta x_{ijds} + \delta z_{jds} + r_{000s} + v_{00ds} + u_{0jds} + e_{ijds},$$

where y_{ijds} is the dichotomous individual match indicator as defined above for student *i* in school *j* in district *d* in state *s*; x_{ijds} is a vector of school-mean centered student covariates, including race indicators; and z_{jds} is a vector of school-level covariates, including proportions for race/ethnicity, gender, and FRPL; school-mean baseline MAP Growth scores; urbanicity; and pandemic vulnerability. r_{000s}, v_{00ds}, u_{0jds}, and e_{ijds} are state, district, school, and residual random effects. We used this model specification because it allowed us to estimate separately and simultaneously within-school differences and between-school differences in test participation (Raudenbush and Bryk, 2002). We interpreted β estimates as average within-school differences based on student characteristics and interpreted δ estimates as average associations between school-mean variables and school-mean match rates.

We used five different model specifications throughout our analysis to test the association between match rates and different student and school characteristics. Model A employs student-level race/ethnicity and baseline MAP Growth reading scores, as well as school-mean race/ethnicity and achievement. This model was used to explore the extent to which matching rates differ depending on whether a student is low- or high-achieving prior to the pandemic and among racial and ethnic groups within schools. Model B regresses matching rates on school-level poverty (as measured by FRPL eligibility), while Model C employs PVI. These models were used to

separately explore associations of matching and poverty and pandemic vulnerability. We did not include school poverty and PVI in a single model because they both capture aspects of the local community's socioeconomic context. Model D uses a combination of FRPL, urbanicity, and the variables from Model A, which helps further disentangle whether differences in match rates by achievement and race/ethnicity occur once we have controlled for poverty. With a similar purpose, Model E uses a combination of PVI, urbanicity, and the variables from Model A.

Study Limitations

There are several limitations to this study that are important to bear in mind when interpreting the results. First, when we were constructing our data set, if an entire school was not present in the follow-up year, we assumed that the school had opted out of testing rather than that all of the students in the school had opted not to participate. We believe that this is a reasonable assumption given what we know about MAP Growth administration. However, we also conducted sensitivity analyses and determined that our regression results were not sensitive to this assumption. Second, there are many operational differences between MAP Growth assessments and state summative assessments that may influence the generalizability of our findings from MAP Growth to the context of state summative assessments. On the one hand, participation in the MAP Growth assessment is not required to the same degree that participation in state summative assessments is required. (For example, there is no Every Student Succeeds Act–mandated 95-percent participation requirement, even in non-COVID-19 school years.) This difference might lead to overstating patterns of missing data in MAP Growth. On the other hand, given the incentives surrounding state exams, even small levels of missing data might be problematic for state summative tests as compared with benchmark and interim tests, including MAP Growth, especially if particular subgroups are systematically missing (see, for example, Figlio and Loeb, 2011). Relatedly, a third limitation of this study is that it was unable to address all state or school assessment contexts. In particular, the participation patterns that we characterize here might not generalize to other assessment contexts, including annual state assessments, such as those required by the Every Student Succeeds Act (e.g., the Smarter Balanced assessments).

Finally, our study focused on one specific issue—changes in test-taking populations and school composition over time—that compromises the comparability of school-level test scores in the pandemic era. However, there are other important questions that threaten the comparability of school-aggregate test scores, including fundamental questions about differences in assessment mode (e.g., remote versus in-person assessment), differences in students' opportunity to learn, and the differential emotional and psychological effects of COVID-19–related stressors. We addressed some of these issues in the discussion section of this report, but, because of data limitations, we could not investigate these issues empirically. Still, we believe that the investigations described in this report help highlight the complexity of interpreting school-level test scores under COVID-19 assessment conditions and help raise awareness of how low or uneven test participation could potentially contaminate aggregate test scores and compromise comparisons among schools or over time.

Results

Student Participation in MAP Growth Assessments Was Lower in 2020–2021 Than in Pre-Pandemic Years

Figure 3 displays test participation rates and corresponding match rates for students in spring 2019 (before the pandemic) and for students in spring 2021 (during the pandemic). The first pair of columns shows the percentage of tested students in the baseline year who were also tested in the follow-up year, even if they moved schools. The second pair of columns shows the percentage of tested students in the baseline year who made expected grade progress, remained enrolled in the same school, and participated in testing in the follow-up year (the longitudinal match rate). During the pandemic period, there was approximately a 10–percentage-point decline in the number of students who participated in testing (from 52 percent to 42 percent), and there was

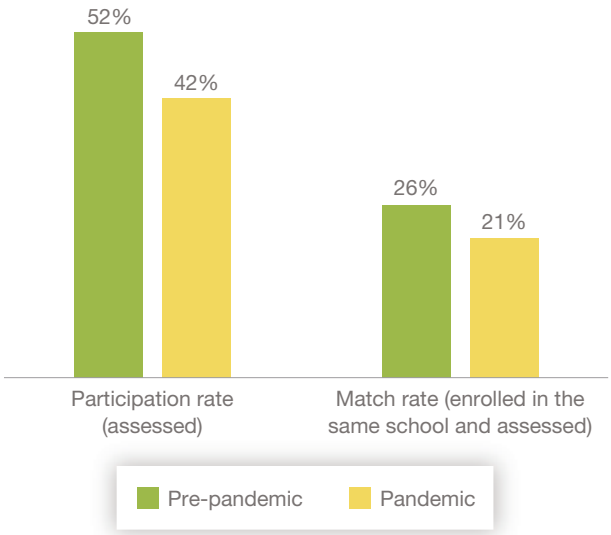
approximately a 5–percentage-point decline in the number of students who were enrolled in the same school and participated in testing (from 26 percent to 21 percent), as compared with the pre-pandemic period. The typical optimal match rate for a school with a configuration of either kindergarten through fifth grade or sixth through eighth grade would be 33 percent, since only students in third grade and sixth grade would have the potential to match, given standard promotional school changes. (As a reminder, data on students in kindergarten through second grade are not included in our analyses.) Of course, in practice, many schools have alternative configurations, schools might not offer MAP Growth assessments in all grades, and newly launched or expanding schools might not have students enrolled across the span of third grade through eighth grade. Because of this, we use this optimal 33-percent benchmark as a heuristic device to guide interpretation. Using this heuristic, we can interpret a 5–percentage-point decline as nearly a 15–percentage-point decline from the expected longitudinal match rate for this sample.

There Was Wide Variability in School-Average Match Rates Within and Among Districts

On average, schools had 26 percent of students match across the pre-pandemic time span. However, in both periods, a substantial number of schools had zero match rates: 13 percent of schools in the pre-pandemic period and 11 percent of schools in the pandemic period. Figure 4 plots the school-level match rate distributions in each period after removing all the schools with zero match rates. There is wide variability in these match rates among schools in both periods. Some schools saw match rates very close to 0 percent, and others had match rates as high as 100 percent.[6] The green density, representing spring 2017 to spring 2019, has a pronounced peak (i.e., mode) around the average match rate of 0.30 (or 30 percent), and the yellow density, representing spring 2019 to spring 2021, has a pronounced peak around the average match rate of 0.26 (or 26 percent). We interpret these differences to mean that, during

FIGURE 3

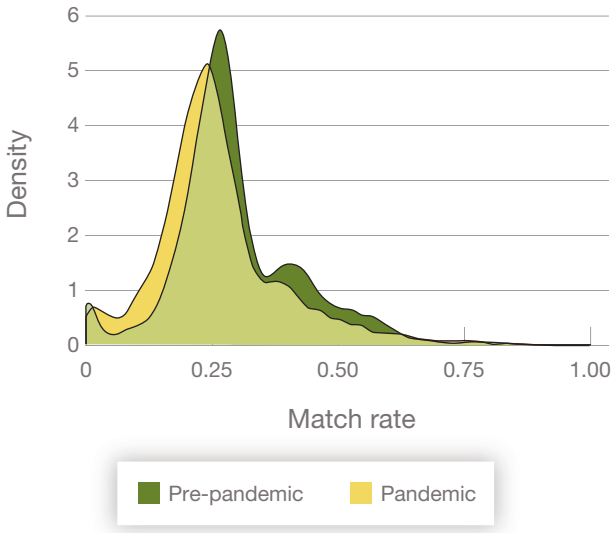
Participation in MAP Growth Reading Tests: Pre-Pandemic and Pandemic Periods



NOTES: *Pre-pandemic* refers to the period from spring 2017 through spring 2019. *Pandemic* refers to the period from spring 2019 through spring 2021.

FIGURE 4

School-Level Match Rate Distributions: Pre-Pandemic and Pandemic Periods



NOTES: Plotted densities are for school-level match rates after removing schools with zero rates. *Pre-pandemic* refers to the period from spring 2017 to spring 2019. *Pandemic* refers to the period from spring 2019 to spring 2021.

the pandemic, overall school-level match rates were typically about 4 percentage points lower than in the pre-pandemic period.

When we looked at the proportion of variance in school match rates within and among districts, we saw that the variance among districts in a state was slightly lower than the variance within a district: Approximately 45 percent of the variance was among districts within states, and approximately 46 percent of the variance was among schools within districts. These percentages were largely similar in the pre-pandemic and pandemic periods, although there was slightly less variance within districts in the pandemic period (Table 5).

To put this variability in match rate into practical terms, the large variance component for districts implies that, within states, some districts had high match rates, and other districts had low match rates. There was an average match rate of 26 percent, so most districts had match rates that ranged from 0 percent to 56 percent. However, there was also a lot of variability within districts; the largest share of the variance in match rates is within districts in both the pre-pandemic and pandemic periods. Some schools within a district might have very high match rates (as high as 100 percent), and other schools within a district might have very low match rates (as low as 0 percent).

TABLE 5

Distributions and Variance Decompositions: School-Level Longitudinal Match Rates

			Variance Decomposition		
Time Period	Mean	Range	State	District	School
Pre-pandemic	30	0–94	8	40	52
Pandemic	26	0–100	8	45	46

NOTES: All numbers are percentages. Percentages might not total 100 because of rounding. Variance decompositions are for schools after removing schools with zero rates. *Pre-pandemic* refers to the period from spring 2017 to spring 2019. *Pandemic* refers to the period from spring 2019 to spring 2021.

There Were Systematic Differences in Match Rates Among Students and Among Communities During the Pandemic Period

To simplify presentation, Table 6 shows the results from a select set of our regression models investigating the extent to which longitudinal match rates differ within and among schools during the pandemic period as a function of student and school characteristics.[7] The first two columns show pre-pandemic results for Models A and B, and the final three columns show pandemic results for Models A, B, and C. For the pre-pandemic period, we do not model the association of PVI with longitudinal match. PVI is a pandemic-specific composite measure that is based on information about COVID-19 infection rates and COVID-19–specific community health practices, including social distancing and testing and, as such, is not well defined in the pre-pandemic period. There were systematic differences in match rates among students and among communities. We highlight key findings from these analyses.

Students of Color Were Generally Less Likely to Be Matched Across Periods Than Their White Peers

In looking at longitudinal match rates across the pandemic period, Black, Asian, Pacific Islander, and Native American students were systematically less likely to be matched than their White peers in the same school. On average, these differences are between 2 and 5 percentage points and are comparable to rates in the pre-pandemic period. Conversely, Hispanic students were about 2 percent more likely to be matched. At the school level, schools serving higher proportions of Black students had systematically lower longitudinal match rates than other schools.

Historically Higher-Achieving Students Were Generally More Likely to Be Matched Across Periods Than Their Peers

In looking at longitudinal matching across the pandemic period, historically higher-achieving students were systematically more likely to be matched than

TABLE 6

Regression Results: Longitudinal Match Rates for Pre-Pandemic and Pandemic Periods

	Pre-Pandemic		Pandemic		
	Model A	**Model B**	**Model A**	**Model B**	**Model C**
Student-level predictors					
Male	0.00	—	0.00***	—	—
	(0.00)		(0.00)		
Black	−0.04***	—	−0.05***	—	—
	(0.00)		(0.00)		
Hispanic	0.02***	—	0.02***	—	—
	(0.00)		(0.00)		
Asian	−0.03***	—	−0.03***	—	—
	(0.00)		(0.00)		
Native American	−0.02***	—	−0.05***	—	—
	(0.01)		(0.01)		
Hawaiian	−0.04***	—	−0.05***	—	—
	(0.01)		(0.01)		
Multiple	−0.03***	—	−0.03***	—	—
	(0.00)		(0.00)		
Other	0.07***	—	0.04***	—	—
	(0.00)		(0.00)		
MAP reading (baseline)	0.03***	—	0.02***	—	—
	(0.00)		(0.00)		
School-level predictors					
Proportion male	−0.04	—	−0.04	—	—
	(0.03)		(0.03)		
Proportion black	−0.18***	—	−0.10***	—	—
	(0.02)		(0.02)		
Proportion Hispanic	0.03*	—	−0.01	—	—
	(0.02)		(0.02)		
Proportion Asian	−0.11**	—	0.01	—	—
	(0.05)		(0.04)		
Proportion Native American	0.04	—	0.01	—	—
	(0.05)		(0.05)		
Proportion Hawaiian	−0.32**	—	0.14	—	—
	(0.17)		(0.16)		

Table 6—Continued

	Pre-Pandemic		Pandemic		
	Model A	Model B	Model A	Model B	Model C
Proportion multiple	−0.12**	—	−0.11*	—	—
	(0.06)		(0.06)		
Proportion other	−0.01	—	0.01	—	—
	(0.02)		(0.02)		
Mean MAP Growth reading score (baseline)	0.08***	—	0.08***	—	—
	(0.01)		(0.01)		
Proportion eligible for FRPL		−0.14***	—	−0.13***	—
		(0.01)		(0.01)	
PVI		—	—	—	−0.14***
					(0.05)
Constant	0.63***	0.70***	0.56***	0.63***	0.63***
	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)
N	705,818	706,631	832,638	833,987	824,909

NOTES: *N* = number of students; * $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$. — = predictor not included in model. *Pandemic* refers to the period from spring 2019 to spring 2021. We also ran models with PVI as an independent predictor in the pre-pandemic period, although those results are not reported. In those models, PVI is not a significant predictor of match rates.

their peers in the same school. On average, a 1–standard deviation increase in MAP Growth scores is associated with a 2-percent increase in match probability over the pandemic period; this result is comparable to the pre-pandemic period result. At the school level, we saw similar trends: Schools serving higher proportions of high-achieving students had systematically higher longitudinal match rates than other schools.

Schools Serving High-Poverty and COVID-19–Vulnerable Communities Had Systematically Lower Match Rates Than Other Schools

Schools that serve higher-poverty and COVID-19–vulnerable communities had systematically lower match rates than other schools across the pandemic period. In particular, Model B shows that schools in which 100 percent of students were eligible for FRPL had approximately 13-percent lower match rates than schools serving communities with 0-percent FRPL eligibility.

Although Some Pandemic Period Participation Trends Are Comparable with Pre-Pandemic Trends, Schools in COVID-19–Vulnerable Communities Had Systematically Lower Match Rates in the Pandemic Period

As shown in Table 6, many of the systematic differences that we saw in the pandemic period based on race and prior academic achievement were also present in the pre-pandemic period: Black, Asian, and Native American students were less likely than their White peers in the same school to be matched over time, and high-achieving students were more likely than their lower-achieving peers in the same school to be matched over time. The extent of these within-school differences in the pre-pandemic period is very similar to the extent of these differences in the pandemic period, which can be seen by visually inspecting the magnitude of the parameter estimates.

At a school level, the patterns are also largely consistent when we compare the pre-pandemic and pandemic periods: Schools enrolling larger shares of Black students or FRPL-eligible students had lower

overall match rates, and schools with higher baseline achievement had higher overall match rates. There are some small differences in the magnitude of these match-rate differences across periods. Schools serving higher proportions of Black students still had lower match rates but to a lesser extent during the pandemic period, whereas schools serving higher proportions of Asian students had considerably higher match rates during the pandemic period.

Given the unique and complex ways in which the pandemic disrupted the lives of students and their families, it is perhaps not surprising that schools in communities that were more vulnerable to COVID-19—those with higher infection rates, higher residential density, higher prevalence of comorbidities, and greater health disparities—had systematically lower match rates: Schools in the most vulnerable communities had match rates that were systematically 14 percent lower than schools in the least vulnerable communities.

Summary

In this report, we used data from NWEA's MAP Growth assessments to investigate one specific issue that may contaminate school-aggregate test scores: changes in school composition that result from low or uneven test participation over time. We investigated whether MAP Growth participation in the pandemic period was lower overall than in pre-pandemic school years and whether match rates varied systematically within and among districts. We also investigated whether there were systematic differences in match rates in assessment depending on student, school, and community characteristics.

Our analyses strongly suggest that student participation in spring assessments in 2020–2021 was both lower relative to recent pre-pandemic years *and* uneven in ways that can contaminate school-aggregate test scores: Fewer students participated in MAP Growth assessments, and there were systematic differences in test participation during the pandemic period. In particular, students of color and schools in economically disadvantaged or COVID-19–affected communities had systematically lower match rates than other schools. These kinds of changes in school populations are likely to bring about immediate changes in school-aggregate measures of academic performance (see, for example, Luyten and de Wolf, 2011). Our analyses also suggest that, although many of the demographic differences in test participation during the pandemic were similar to pre-pandemic patterns, schools in COVID-19–vulnerable communities had systematically lower match rates in the pandemic period.

Finally, the fact that there was considerable variability in test participation within and among schools should not be overlooked. In some schools, participation rates were quite high. In others, participation rates were quite low. The results from our regression-based analyses describe average within- and among-school differences in test participation, and these averages might or might not accurately describe any specific school or district. For example, there might be geographical interactions that alter patterns of test participation among subgroups: It might be that, in some regions or school districts, Black students were less likely to participate in testing, while, in other regions, White students were less likely to participate.

Our analyses strongly suggest that student participation in spring assessments in 2020–2021 was both lower relative to recent pre-pandemic years *and* uneven in ways that can contaminate school-aggregate test scores.

It is important to keep in mind that this study focused on only one issue that might compromise school-aggregate test score interpretation. Here, we briefly discuss two additional issues that school systems will want to examine to better understand the ways in which comparisons among schools or over time might be compromised: differences in assessment mode (i.e., remote versus in-person assessment) and differences in students' opportunity to learn.

Generally speaking, in addition to assuming stable test-taking populations, school-aggregate test score comparisons assume that all individuals took the same test under the same or similar testing conditions (DePascale and Gong, 2020). Although variations in testing conditions (e.g., one student might be having an off day, another student might be dealing with a broken pencil or a broken air conditioner, another student might encounter a set of test questions that aligns perfectly with the material they know best) are generally treated as idiosyncratic, the pandemic introduces the potential for differences in testing conditions to vary systematically among schools. In particular, some schools were more likely to administer tests remotely than others. A national survey of public school systems in winter 2021 reported that 18 percent of districts offered instruction that was fully in person and 10 percent offered instruction that was mostly or fully remote (Hodgman et al., 2021). Given that U.S. ED guidance to states indicated that students should not be brought to school for the sole purpose of test administration,[8] it is likely that there were similar variations in the percentages of students who participated in assessment remotely and in person in spring 2021.

To illustrate this point, imagine that two schools administered the same test, but students in one school took the test remotely at home and students in the other school took the test in person at school. Aggregate score differences might reflect real differences in school-level achievement but also might reflect other factors about the testing context, including technological limitations; the availability of parent support; and access to clean, quiet, and distraction-free spaces in which to take the test. Alternatively, some students might feel more comfortable when testing outside a school setting, and standardized testing procedures might interact with personal characteristics in ways that systematically affect test performance in undesirable ways (Sireci, 2020).

There is another aspect that threatens the comparability of aggregate test scores, even if the same test is given at all schools and if administration conditions are controlled to the extent possible: opportunity to learn, which includes content coverage, content exposure, content emphasis, and the quality of instructional delivery (McDonnell, 1995; Stevens and Grymes, 1993). Opportunity to learn has long been considered an important predictor of learning outcomes (Martínez, 2012; Wang and Goldschmidt, 1999), and content exposure is associated with student academic achievement (Goodman, Miller, and West-Olatunji, 2012; Lavy, 2015). Opportunity to learn plays an important role in determining whether test scores can be used fairly and accurately to monitor school progress and performance (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 2014). COVID-19 presents special challenges for disentangling differences in opportunity to learn from differences in test administration because differences in test administration mode (remote or in person) are likely confounded

Although variations in testing conditions are generally treated as idiosyncratic, the pandemic introduces the potential for differences in testing conditions to vary systematically among schools.

The potential impacts of differences in opportunity to learn during COVID-19 might be substantial because school closures had substantial impacts on both content exposure and the quality of instructional delivery.

with changes in instructional mode. The potential impacts of differences in opportunity to learn during COVID-19 might be substantial because school closures had substantial impacts on both content exposure and the quality of instructional delivery. There is evidence that instructional effects may have been the largest in the most-disadvantaged schools and that remote learning may have had the most adverse impacts for the most-vulnerable students (e.g., Haderlein et al., 2021; Hodgman et al., 2021; Kaufman and Diliberti, 2021).

Discussion

As school systems respond to the COVID-19 pandemic by developing policies and practices that focus on restart and recovery, many systems will almost certainly rely on school-aggregate scores from state summative, interim, and benchmark assessments to identify schools where students have responded well to COVID-19 disruption, restart, and recovery, as well as schools where students have been disproportionately affected. However, using school-aggregate test scores for such purposes relies on the assumption that differences in aggregate test scores from spring 2019 to spring 2021 can be accurately interpreted as representing real and meaningful differences in school progress and performance during the pandemic. Central to evaluating the accuracy of school-aggregate test scores is understanding the extent to which the scores have been influenced by low and uneven test participation, systematic differences in test administration mode, and systematic differences in opportunity to learn among school contexts during the pandemic. This study offers three insights relative to test participation that might provide guidance to school systems as they report school-level test score information from spring 2021 and implement restart and recovery plans.

Comparing Spring 2021 School-Aggregate Test Scores with Those from Spring 2019 Might Misrepresent School Progress or Pandemic Impacts

Our analysis of students' match rates shows that students in communities that were more likely to be affected by the pandemic were much less likely to be matched across the 2019 to 2021 school years. Specifically, students in the most affected communities were 15 percent less likely to be matched than their peers in the least affected communities. If students and communities that were likely to be more affected by the pandemic did not participate in testing in spring 2021, comparisons with spring 2019 might *overstate* school progress. This, in turn, suggests that the identification of schools in which students have been disproportionately affected by COVID-19 might *underestimate* the extent of the problem if educators and policymakers do not take test participation into consideration as an important contextual factor for interpreting school-aggregate test scores (Barnum, 2021).

This differential missingness will also make it difficult to use test scores to inform resource allocation (e.g., states' and districts' distribution of Elementary and Secondary School Emergency Relief funds) and to evaluate interventions focused on restart and

recovery. For this reason, as a general rule, the findings of this study suggest that comparisons of school-aggregate MAP Growth scores over time, especially comparisons that focus on COVID-19 versus pre-COVID-19 school years, should be avoided. Such a recommendation is consistent with statements made by several states, on their websites or in accompanying technical documentation, that 2019 and 2021 school-level scores from state summative tests should not be compared because of participation issues.

However, we acknowledge that the aggregate trends presented in our analyses might not apply equally to all schools and districts. Indeed, we found wide variability in test-taking patterns among districts and schools. This strongly suggests that school systems should conduct their own local analyses of longitudinal match rates for schools overall and by subgroup to better understand patterns of low or uneven test participation over time and should make local determinations about the extent to which school-level comparisons over time might be compromised by changes in school composition. In systems in which participation was high and consistent over time, comparisons might still be useful for identifying school needs and allocating resources.

The findings of this study suggest that comparisons of school-aggregate MAP Growth scores over time, especially comparisons that focus on COVID-19 versus pre-COVID-19 school years, should be avoided.

Comparing Spring 2021 Aggregates Among Schools Might Misrepresent the Relative Performance of Schools

Given that participation in assessment was lower in spring 2021, comparisons among schools or districts in spring 2021 are likely compromised, and we caution against comparing aggregate test scores among schools without setting such comparisons in the appropriate context of test participation. Participation differences might contaminate school comparisons even as school systems emerge from the pandemic and schooling returns to a sense of normalcy. We recognize that school system leaders have access to a variety of data sources to inform their understanding of how COVID-19 affected students and schools in the 2020–2021 school year, and officials should, to the extent possible, base appraisals of school performance, progress, and COVID-19 response on multiple sources of information, such as benchmark and formative assessments, locally determined measures of COVID-19 impacts, and data on school or community context. As above, this recommendation is made with an acknowledgment that test participation varied among schools and districts; school systems should conduct their own test participation analyses to determine the extent to which comparisons among schools might be compromised by low or uneven test participation. Such analyses might involve having principals or other school leaders conduct school-specific investigations of test participation to better understand how many students—and which ones—participated in spring 2021 testing (Gewertz, 2021). In systems in which participation was high among student subgroups, comparisons might still be useful for identifying school needs and allocating resources.[9]

Information About Low and Differential Participation in Spring 2021 Testing Should Be Transparently Reported Along with Publicly Released Data

Although this study focused on test participation in the context of the NWEA MAP Growth tests, the empirical results of this study largely support recommendations from researchers and policymakers that

When systems report school-aggregate test scores, they should consider how the pandemic has affected their communities and the ways in which they can address this in the presentation of school-average performance.

spring 2021 school-level reporting of state summative test scores be accompanied by clear information about historical and current test participation (Ho, 2021; U.S. ED, 2021b). Providing this information allows the public to appraise the extent to which test results are representative of school populations and to evaluate the extent to which comparisons among schools or over time may be compromised by changes in school composition. We caution that, although such reporting practices are important as a foundational practice, many administrative data sets contain only a limited number of variables that describe student or school characteristics, and these characteristics might not fully capture the myriad ways in which the pandemic affected students and communities. It is thus important for systems to communicate that reporting of broad averages, although useful, might misspecify community or school trends or other important details about how students, communities, and schools have fared over the past three school years. When systems report school-aggregate test scores, they should consider how the pandemic has affected their communities and the ways in which they can address this in the presentation of school-average performance.

Acknowledgments

The authors are grateful to the expert panelists who provided insight and guidance on this report: June Ahn, Sharon Bi, Dan Goldhaber, Betheny Gross, Laura Hamilton, Matthew Raimondi, Kevin Schaaf, Jessaca Spybrook, and Katharine Strunk. This document benefited substantively from feedback from Morgan Polikoff and Bonnie Ghosh-Dastidar. Emily Ward provided expert editing. Any flaws that remain are solely the authors' responsibility.

Notes

[1] However, by May 2021, only 1 percent of U.S. school districts had fully remote instruction (Ferren, 2021).

[2] Many states note that low and uneven participation during spring 2021 compromises comparisons among schools or over time, and several states caution against making comparisons without fully understanding the assessment context and test participation rates.

[3] State policy is evolving rapidly in response to the pandemic, and it is difficult to locate accurate public documentation of current state plans. We corroborated these claims through conversations with accountability directors in five states.

[4] Similar analyses were conducted for math assessments, but, because there were very few students who were assessed in one subject but not the other, results did not differ in substance or interpretation. Math results are available upon request from the authors.

[5] Note that N_{CD} and N_C are school-level counts. For ease of exposition, we omit the school-level subscript *j*.

[6] Schools with configurations of kindergarten through fifth grade and sixth through eighth grade have an expected optimal match rate of 33 percent. Other school and testing configurations can result in higher match rates. In our sample, approximately 2 percent of students are enrolled in schools with alternative grade configurations. Optimal match rates might also be higher for new or expanding schools. For example, a school that enrolled and tested only third-graders in 2019 could have a 100-percent match rate if all fifth-graders were assessed in 2021.

[7] Results for the two additional models (D and E) were consistent with the results presented here, so, for ease of presentation, we have omitted those results from this report. Complete results are available from the authors upon request.

[8] See Rosenblum, 2021.

[9] Some test designers have cautioned that participation rates below 50 percent would seriously compromise comparability, even if there were no systematic differences in student participation (Gewertz, 2021).

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing*, 2014 ed., Washington, D.C.: American Educational Research Association, 2014.

Barnum, Matt, "This Year's State Test Results Will Be Tough to Make Sense of, Experts Warn," Chalkbeat, February 24, 2021.

Beaver, Jessica K., and Elliot H. Weinbaum, "State Test Data and School Improvement Efforts," *Educational Policy*, Vol. 29, No. 3, 2015, pp. 478–503.

Beaver, Jessica K., Lucas Westmaas, and John Sludden, *The Potential Effects of Opting Out of State Tests in Pennsylvania*, Philadelphia, Pa.: Research for Action, 2014.

Bouzaghrane, Mohamed Amine, Hassan Obeid, Madeleine Parker, Meiqing Li, Drake Hayes, Minnie Chen, Karen Trapenberg Frick, Daniel Rodríguez, Joan Walker, Raja Sengupta, Daniel G. Chatman, *Research Brief: The Changing Impacts of the COVID-19 Pandemic on Individuals and Households in the U.S.*, Berkeley, Calif.: University of California, Berkeley, January 19, 2021.

Bruno, Paul, and Dan Goldhaber, *Reflections on What Pandemic-Related State Test Waiver Requests Suggest About the Priorities for the Use of Tests*, Arlington, Va.: National Center for Analysis of Longitudinal Data in Education Research, CALDER Policy Brief No. 26-0721, 2021.

Bryant, Jeff, "The Fight Against Standardized Testing Is More Diverse Than You Think," *Progressive Magazine*, March 4, 2016.

California Department of Education, "State Board of Education Approves Additional Testing Flexibility for Schools This Spring," press release, Sacramento, Calif., March 16, 2021.

Caprariello, Alex, "New STAAR Data Reveals Texas Students Slipped Significantly in Reading and Math," KXAN, last updated June 29, 2021.

Center on Reinventing Public Education, "State Guidance on School Reopenings amid the Pandemic," webpage, 2021.

Centers for Disease Control and Prevention, "CDC Offers Health Tips for Back to School During COVID-19," press release, Atlanta, Ga., August 23, 2021.

Colorado Department of Education, "Spring 2021 CMAS Data and Results," webpage, 2021. As of October 1, 2021: https://www.cde.state.co.us/assessment/cmas-dataandresults-2021

Council of Chief State School Officers, "CCSSO Releases Statement on Assessments in the Next Academic Year," press release, Washington, D.C., July 20, 2020.

Cremata, Edward, *Who's Missing? Exploring the Magnitude and Impact of Student Opt-Outs on School Accountability Systems*, Stanford, Calif.: Policy Analysis for California Education, 2019.

DePascale, Charles, and Brian Gong, "Comparability of Individual Students' Scores on the 'Same Test,'" in Amy I. Berman, Edward H. Haertel, and James W. Pellegrino, eds., *Comparability of Large-Scale Educational Assessments: Issues and Recommendations*, Washington, D.C.: National Academy of Education, 2020, pp. 25–48.

Digital Promise, *Learning in the 21st Century: How the American Public, Parents, and Teachers View K–12 Teaching and Learning in the Pandemic*, Washington, D.C., 2021.

Dougherty, Chrys, *Use of Data to Support Teaching and Learning: A Case Study of Two School Districts*, Iowa City, Ia.: ACT, 2015.

Education Analytics, *Evaluating the Impact of COVID-19-Related Interventions: Key Considerations and Available Approaches*, Madison, Wisc.: September 2021.

Education Policy Innovation Collaborative, *Michigan's 2020–21 Benchmark Assessments*, East Lansing, Mich., 2021.

Ferren, Megan, *Remote Learning and School Reopenings: What Worked and What Didn't*, Washington, D.C.: Center for American Progress, 2021.

Figlio, David, and Susanna Loeb, "School Accountability," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 3, Amsterdam: North-Holland, 2011, pp. 383–421.

García, Emma, and Elaine Weiss, *COVID-19 and Student Performance, Equity, and U.S. Education Policy: Lessons from Pre-Pandemic Research to Inform Relief, Recovery, and Rebuilding*, Washington, D.C.: Economic Policy Institute, 2020.

Gewertz, Catherine, "It's Official: All States Have Been Excused from Statewide Testing This Year," *Education Week*, April 2, 2020.

———, "State Test Results Are In. Are They Useless?" *Education Week*, October 21, 2021.

Goodman, Rachael D., M. David Miller, and Cirecie A. West-Olatunji, "Traumatic Stress, Socioeconomic Status, and Academic Achievement Among Primary School Students," *Psychological Trauma: Theory, Research, Practice, and Policy*, Vol. 4, No. 3, 2012, pp. 252–259.

Haderlein, Shira K., Anna Rosefsky Saavedra, Morgan S. Polikoff, Daniel Silver, Amie Rapaport, and Marshall Garland, "Disparities in Educational Access in the Time of COVID: Evidence from a Nationally Representative Panel of American Families," *AERA Open*, Vol. 7, No. 1, January–December 2021, pp. 1–21.

Hamilton, Laura S., and Kadriye Ercikan, "COVID-19 and U.S. Schools: Using Data to Understand and Mitigate Inequities in Instruction and Learning," in Fernando M. Reimers, eds., *Primary and Secondary Education During Covid-19*, Cham, Switzerland: Springer, 2022, pp. 327–351.

Hamilton, Laura, Richard Halverson, Sharnell S. Jackson, Ellen Mandinach, Jonathan A. Supovitz, and Jeffrey C. Wayman, *Using Student Achievement Data to Support Instructional Decision Making*, Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, NCEE 2009-4067, 2009.

Herold, Benjamin, "The Decline of Hybrid Learning for This School Year, in 4 Charts," *Education Week*, September 27, 2021.

Ho, Andrew, "Three-Test Score Metrics That All States Should Report in the COVID-19-Affected Spring of 2021," draft memo, February 26, 2021.

Hodgman, Sarah, Jordan Rickles, Joanne Carminucci, and Mike Garet, *District Approaches to Instruction in 2020–21: Differences in Instructional Modes and Instruction Time Across Contexts*, Arlington, Va.: American Institutes for Research, 2021.

Institute of Education Sciences, "Monthly School Survey Dashboard," webpage, undated. As of November 2, 2021: https://ies.ed.gov/schoolsurvey/

Jacobson, Linda, "Pandemic Testing Gaps Complicate Ability to Pinpoint Struggling Schools at a Time When Students Need Extra Help, School Leaders Say," The 74, September 16, 2021a.

———, "As the Pandemic Set In, Charter Schools Saw Their Highest Enrollment Growth Since 2015, 42-State Analysis Shows," The 74, September 22, 2021b.

Johnson, Angela, and Megan Kuhfeld, *Fall 2019 to Fall 2020 MAP Growth Attrition Analysis*, Portland, Oreg.: NWEA, 2020.

Kaufman, Julia H., and Melissa Kay Diliberti, *Divergent and Inequitable Teaching and Learning Pathways During (and Perhaps Beyond) the Pandemic: Key Findings from the American Educator Panels Spring 2021 COVID-19 Surveys*, Santa Monica, Calif.: RAND Corporation, RR-A168-6, 2021. As of November 2, 2021: https://www.rand.org/pubs/research_reports/RRA168-6.html

Keng, Leslie, and Scott Marion, "Comparability of Aggregated Group Scores on the 'Same Test,'" in Amy I. Berman, Edward H. Haertel, and James W. Pellegrino, eds., *Comparability of Large-Scale Educational Assessments: Issues and Recommendations*, Washington, D.C.: National Academy of Education, 2020, pp. 49–74.

Kurtz, Holly, "In-Person Learning Expands, Student Absences Up, Teachers Work Longer, Survey Shows," *Education Week*, October 15, 2020.

Lavy, Victor, "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries," *Economic Journal*, Vol. 125, No. 588, November 2015, pp. F397–F424.

Lleras, Christy, and Mary McKillip, "When Children Move: Behavior and Achievement Outcomes During Elementary School," *Journal of Educational Research*, Vol. 110, No. 2, 2017, pp. 177–187.

Luyten, Hans, and Inge de Wolf, "Changes in Student Populations and Average Test Scores of Dutch Primary Schools," *School Effectiveness and School Improvement*, Vol. 22, No. 4, 2011, pp. 439–460.

Mahnken, Kevin, "Research Shows Changing Schools Can Make or Break a Student, But the Wave of Post-COVID Mobility May Challenge the Systems in Ways We've Never Seen," The 74, December 15, 2021.

Martínez, José Felipe, "Consequences of Omitting the Classroom in Multilevel Models of Schooling: An Illustration Using Opportunity to Learn and Reading Achievement," *School Effectiveness and School Improvement*, Vol. 23, No. 3, 2012, pp. 305–326.

Marvel, Skylar W., John S. House, Matthew Wheeler, Kuncheng Song, Yi-Hui Zhou, Fred A. Wright, Weihsueh A. Chiu, Ivan Rusyn, Alison Motsinger-Reif, and David M. Reif, "The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: Monitoring County-Level Vulnerability Using Visualization, Statistical Modeling, and Machine Learning," *Environmental Health Perspectives*, Vol. 129, No. 1, 2021.

McDonnell, Lorraine M., "Opportunity to Learn as a Research Concept and a Policy Instrument," *Educational Evaluation and Policy Analysis*, Vol. 17, No. 3, Fall 1995, pp. 305–322.

McEachin, Andrew, Thurston Domina, and Andrew Penner, "Heterogeneous Effects of Early Algebra Across California Middle Schools," *Journal of Policy Analysis and Management*, Vol. 39, No. 3, Summer 2020, pp. 772–800.

Means, Barbara, Christine Padilla, and Larry Gallagher, *Use of Education Data at the Local Level: From Accountability to Instructional Improvement*, Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, 2010.

Michigan Department of Education, *Spring 2021 Interpretive Guide to M-Step Reports*, September 2021.

National Center for Education Statistics, "CCD Data Files," webpage, undated. As of October 1, 2021: https://nces.ed.gov/ccd/files.asp

———, "New Data Reveal Public School Enrollment Decreased 3 Percent in 2020–21 School Year," *NCES Blog*, July 26, 2021.

National Institute of Environmental Health Sciences, homepage, undated a. As of December 2, 2021: https://www.niehs.nih.gov/

———, "COVID-19 Pandemic Vulnerability Index (PVI)," webpage, undated b. As of October 1, 2021: https://covid19pvi.niehs.nih.gov/

New Jersey Department of Education, Office of Assessments, "Start Strong Fall 2021 Administration," presentation slides, June 2021.

NWEA, "MAP Growth Instructional Area," webpage, undated. As of October 1, 2021: https://cdn.nwea.org/state-information/index.html#/

O'Donnell, Patrick, "The Concerning Case of Cleveland's No-Show Students: More Than 8,000 Kids Are Missing from City's Online Classes as Absenteeism Rates Double," The 74, December 16, 2020.

Peele, Holly, and Maya Riser-Kositsky, "Map: Coronavirus and School Closures in 2019–2020," *Education Week*, last updated October 13, 2021.

Polikoff, Morgan, *On Growth Models, Time for California to Show Some Improvement*, Stanford, Calif.: Policy Analysis for California Education, 2019.

Public Law 116-136, Coronavirus Aid, Relief, and Economic Security Act, March 27, 2020.

Public Law 116-260, Coronavirus Response and Relief Supplemental Appropriations Act, December 27, 2020.

Public Law 117-2, American Rescue Plan Act of 2021, March 11, 2021.

Raudenbush, Stephen W., and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, Vol. 1, Thousand Oaks, Calif.: SAGE Publications, Inc., 2002.

Rosenblum, Ian, memorandum to chief state school officer, Washington, D.C.: U.S. Department of Education, Office of Elementary and Secondary Education, February 22, 2021.

Schweig, Jonathan, "The Opt-Out Reckoning: An Ever-Growing Call to Opt Out of Standardized Tests Is Prompting Serious Questions in Education," *U.S. News & World Report*, May 9, 2016.

Schweig, Jonathan, Andrew McEachin, Megan Kuhfeld, Louis T. Mariano, and Melissa Kay Diliberti, *Adapting Course Placement Processes in Response to COVID-19 Disruptions: Guidance for Schools and Districts*, Santa Monica, Calif.: RAND Corporation, RR-A1037-1, 2021. As of December 2, 2021: https://www.rand.org/pubs/research_reports/RRA1037-1.html

Sireci, Stephen G., "Standardization and UNDERSTANDardization in Educational Assessment," *Educational Measurement: Issues and Practice*, Vol. 39, No. 3, Fall 2020, pp. 100–105.

Stevens, Floraline I., and John Grymes, *Opportunity to Learn: Issues of Equity for Poor and Minority Students*, Washington, D.C.: National Center for Education Statistics, 1993.

Stokes, Kyle, "LAUSD Won't Participate in Statewide Testing for Second Straight Year," LAist, April 15, 2021.

U.S. Department of Education, "U.S. Department of Education Releases Guidance to States on Assessing Student Learning During the Pandemic," press release, Washington, D.C., February 22, 2021a.

———, *Strategies for Using American Rescue Plan Funding to Address the Impact of Lost Instructional Time*, Washington D.C.: August 2021b.

U.S. Department of Education, Office for Civil Rights, *Education in a Pandemic: The Disparate Impacts of COVID-19 on America's Students*, Washington, D.C.: 2021.

U.S. Department of Education, Office of Elementary and Secondary Education, "State Requests for Waivers of ESEA Provisions for SSA-Administered Programs," webpage, last modified June 29, 2021. As of September 28, 2021: https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/essa-state-plans-assessment-waivers/

U.S. ED—*See* U.S. Department of Education.

Wang, Jia, and Pete Goldschmidt, "Opportunity to Learn, Language Proficiency, and Immigrant Status Effects on Mathematics Achievement," *Journal of Educational Research*, Vol. 93, No. 2, November–December 1999, pp. 101–111.

Welsh, Richard O., "School Hopscotch: A Comprehensive Review of K–12 Student Mobility in the United States," *Review of Educational Research*, Vol. 87, No. 3, June 2017, pp. 475–511.

West Virginia Department of Education, *2021 West Virginia Statewide Summative Assessment Results*, Charleston, W.Va., 2021.

Wise, S. L., M. R. Kuhfeld, and J. Cronin, *Assessment in the Time of COVID-19: Engagement During Low-Stakes Remote Testing*, paper presented at the annual meeting of the National Council on Measurement in Education, virtual presentation, June 2021.

About This Report

In February 2021, the U.S. Department of Education announced that states would be encouraged to administer statewide summative assessments at the end of the 2020–2021 school year to identify student needs and to target resources to school communities that have been disproportionately affected by the coronavirus disease 2019 (COVID-19) pandemic. Although many states administered the same tests that had been in use prior to the pandemic, others implemented policies that placed an increased emphasis on the use of interim or benchmark assessments like NWEA's Measures of Academic Progress (MAP) Growth assessments to monitor and promote restart and recovery.

Although individual student scores are fundamental for appraising student progress, school-aggregate test scores are also an essential source of information and play a key role in system plans to address unfinished learning and opportunity gaps at the classroom, school, district, and state levels. However, using aggregate test scores to monitor school performance relies on the assumption that differences in aggregate test scores can be accurately interpreted as representing real and meaningful differences in school progress and performance. In the context of the pandemic, there are several issues that complicate this interpretation and thus compromise the comparability of aggregate-level test scores, both over time and among schools and districts.

In this study, RAND researchers used MAP Growth data from NWEA's Growth Research Database to investigate one issue that may contaminate utilization of COVID-19–era school-aggregate scores: changes in test-taking populations and school composition over time. This report is the second of three that examine the impacts of COVID-19–related assessment disruptions on school and district processes. The first report, *Adapting Course Placement Processes in Response to COVID-19 Disruptions: Guidance for Schools and Districts*, compares three strategies to estimate missing test scores and help with course placement decisions (Schweig et al., 2021). A future report will address strategic decisionmaking for research and evaluation. The purpose of this study is to illustrate how changes in test-taking populations during the pandemic can influence the accuracy of determinations about school recovery that are made based on benchmark assessments like MAP Growth. The study's ultimate goal is to highlight important considerations that should guide policy and practice around school-aggregate test score use.

RAND Education and Labor

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. The research reported here was sponsored by the Institute of Education Sciences, U.S. Department of Education, through grant R305U200006 to RAND. The opinions expressed in this report are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to jschweig@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

Research Integrity

Our mission to help improve policy and decisionmaking through research and analysis is enabled through our core values of quality and objectivity and our unwavering commitment to the highest level of integrity and ethical behavior. To help ensure our research and analysis are rigorous, objective, and nonpartisan, we subject our research publications to a robust and exacting quality-assurance process; avoid both the appearance and reality of financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursue transparency in our research engagements through our commitment to the open publication of our research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. For more information, visit www.rand.org/about/principles.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

For more information on this publication, visit www.rand.org/t/RRA1037-2.

© 2022 RAND Corporation

www.rand.org