Article

Journal of Early Adolescence
2026, Vol. 46(2) 278–305

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02724316251317596
journals.sagepub.com/home/jea

Combined Schools as a (Non) Protective Factor During the COVID-19 Pandemic: Exploring Middle School Transitions Under Typical versus Atypical School Conditions

Melissa K. Diliberti[1], Jonathan D. Schweig[1], Louis T. Mariano[1], Megan Kuhfeld[2], and Andrew McEachin[3]

Abstract

Combined schools spanning the elementary and middle grades (like K–8 schools) can protect students from harms associated with structural transitions by offering a stable school environment at a time when youth are undergoing many physical, psychological, and social changes. Using math and reading MAP Growth assessment scores from 20,000 students in parallel prepandemic and pandemic event studies, we investigate the extent to which the protective benefits historically offered by combined schools prepandemic held during the COVID-19 pandemic—a period of prolong disruption to normal schooling conditions. We find sixth and seventh graders enrolled in combined schools before the pandemic had scores 0.12 to 0.16 standard

[1]RAND Corporation, Santa Monica, CA, USA
[2]NWEA, Portland, OR, USA
[3]ETS Research Institute, Princeton, NJ

Corresponding Author:
Melissa K. Diliberti, RAND Corporation, 4570 Fifth Ave #600, Pittsburgh, PA 15232, USA.
Email: mdiliber@rand.org

deviation (SD) units higher than their peers in standalone middle schools. That advantage diminished during the pandemic; middle grade youth in combined schools in 2020–2021 and 2021–2022 had math and reading scores that were only 0.04 to 0.08 SDs higher than their transitional counterparts. During the pandemic, school configuration mattered more for students' reading, but not their math, achievement if they were attending school in-person than if they were attending school remotely. Our findings raise questions about the mechanisms through which combined schools offer academic benefits under both typical and atypical schooling conditions.

Keywords
middle school transitions, COVID-19 pandemic, student achievement, K–8 schools, school disruptions

How best to configure school grade spans for students' learning, especially at the middle grades, has been a focus for decades (see, e.g., Juvonen, 2004; Mergendoller, 1993). The shock of the structural middle school transition—or students' forced transition to new schools brought on by terminating elementary school grades—negatively impacts students in many ways. Early work found this structural transition to be associated with many negative academic, social, and emotional outcomes such as decreased student achievement, increased dropout rates, increased absenteeism, increased discipline problems, decreased motivation, and lower self-esteem (see, e.g., Alspaugh, 1998, 2000; Barber & Olsen, 2004; Byrnes & Ruby, 2007; Eccles et al., 1991; Weiss & Kipnes, 2006). Importantly, vulnerable student groups—such as minority students, historically low-achieving students, and low-income students—likely feel the negative impacts of these middle school transitions most acutely (see, e.g., Atteberry et al., 2022; Hong et al., 2018; Schwartz et al., 2011; Schwerdt & West, 2013).

Combined schools that bridge the elementary and middle grades like K–8 or K–12 schools (sometimes called "elemiddle" schools) are a policy alternative to standalone middle schools that have become more common over the last two decades (Juvonen, 2004). Combined schools are hypothesized to protect students against the academic declines historically associated with making a structural middle school transition by holding at least one factor constant (the school environment) during a time when youth are undergoing many other personal, physical, psychological, and social changes. Said another way, combined schools seek to minimize changes to the school environment for middle grade students, with the hope that a more stable school environment will positively impact students' academic achievement. Prior work suggests that combined schools can be an effective intervention to support student achievement in the middle grades. Multiple studies have found that structural middle school transitions negatively affect achievement and that middle grade students in combined schools typically outperform their peers who make structural transitions (Atteberry et al., 2022; Byrnes & Ruby, 2007; Dhuey, 2013; Hong et al., 2018; Rockoff & Lockwood, 2010; Schwartz et al., 2011; Schwerdt & West, 2013).

However, it is unclear whether the protective benefits historically offered by combined schools held during the COVID-19 pandemic, a time when normal schooling conditions were widely disrupted. We hypothesized two opposing scenarios for how combined schools may have functioned in this atypical context. On the one hand, we might expect the achievement of youth who had to undergo a structural middle school transition during the pandemic to suffer to a greater degree than normal, given that these youth were dealing with additional stressors above and beyond those normally experienced. If these transitional students experienced larger-than-normal academic declines while the protective benefits of combined schools held, we would expect to observe a *greater* divergence between students in combined schools and those who transitioned than observed under typical schooling conditions. On the other hand, the theoretical benefits of the combined school are predicated upon assumptions of continuity and stability in the school environment—assumptions that were compromised since half of U.S. districts were still providing mostly remote or hybrid instruction until well into the spring of the 2020–2021 school year (American Enterprise Institute, 2021). Thus, the mechanisms normally disproportionately present in combined schools that help create a stable learning environment—things like continuous teacher-student relationships and established norms and routines—might have diminished during the pandemic. If this is true, then we would expect to observe a *smaller* divergence than normal in students' academic achievement in the middle grades. Thus, it is not immediately obvious whether combined schools offer academic protections during atypical schooling conditions. To find out, we conducted parallel prepandemic and pandemic event studies in which we compared the reading and math achievement of sixth and seventh graders (when youth were around 10- or 11-years-old) who remained in a combined school with those who made a structural middle school transition.

Background

Schools as a Key Setting for Students' Adolescent and Academic Development

Youth experience many physical, psychological, and social changes as they enter adolescence and begin to form their own identity (Sebastian et al., 2008). Given the amount of time youth spend in school, the school environment is a key context in which to support these developmental changes (Eccles & Roeser, 2011). Eccles and Midgley (1988, 1993) describe this as "stage-environment fit," or the idea that school environments should be structured to appropriately align with the developmental stage of their students. And yet middle schools tend to use practices (e.g., larger school environments, less personal teacher interactions, sorting students into classes by ability level, departmentalized instruction) incongruent with students' developmental stage (Eccles & Midgley, 1988; Eccles & Roeser, 2011).

The declines in self-esteem and intrinsic motivation, feelings of ill preparedness, lower grades, and lower achievement that arise during adolescence are well documented (see, e.g., Alspaugh, 1998; Eccles et al., 1991; Harter, 1981; Seidman et al., 1994). Over the past three decades, researchers have increasingly focused on the forced middle school transition to explain poor outcomes in the middle school grades, particularly for vulnerable student subgroups (see, e.g., Atteberry et al., 2022; Byrnes & Ruby, 2007; Rockoff & Lockwood, 2010). Addressing the mismatch between middle school structures and students' developmental stage to better support students' social and academic development has been a key aspect of middle school reform efforts. These efforts have emphasized the need to build middle school environments that promote continuity and social connectedness—or environments that perhaps more closely mirror those in elementary schools (Carnegie Council on Adolescent Development, 1989; Juvonen, 2007). Policymakers and practitioners hope that by reforming middle school environments, they will improve academic outcomes in the middle grades.

The Mechanisms Through Which Combined Schools Work to Bolster Achievement

The literature offers several hypotheses of the mechanisms through which combined schools seek to bolster students' academic achievement. For one, the combined school structure may offer students greater continuity in their learning environment. Eccles and Midgley (1988) describe the classroom changes that students usually encounter for the first time in standalone middle schools such as less personal teacher-student interactions, stricter grading, ability grouping, and other changes emphasizing social comparisons. Middle grade classrooms in combined schools may use instructional practices that more closely mirror their elementary counterparts and/or students' shock at encountering these instructional changes might be lessened in a familiar school environment. Second, the large enrollment sizes of standalone middle schools have repeatedly been linked to negative student outcomes, perhaps in part because there are fewer opportunities for students to receive personalized attention (Byrnes & Ruby, 2007; Eccles et al., 1991; Weiss & Kipnes, 2006). As speculated by Offenberg (2001), smaller cohort sizes in combined schools likely also translate into fewer teachers per grade. In sum, a smaller environment in combined schools likely allows for more and more personalized student-student, teacher-student, and teacher-parent interactions, potentially fostering a greater sense of community as well as less transitory relationships. Third, middle grade youth in combined schools perceive more positive school climates and experience less bullying than their peers in standalone middle schools (Anderman & Kimweli, 1997; Malone et al., 2017; Weiss & Kipnes, 2006). Youth in combined schools do not experience being "bottom dog" after being "top dog" in their last year of elementary school, a phenomenon associated with increased bullying and fewer feelings of belonging (Schwartz et al., 2016). Thus, the more welcoming school climate in combined schools may, in turn, lead to improved academic outcomes.

What Changed During the COVID-19 Pandemic

For those entering the middle school grades in 2020–2021, the COVID-19 pandemic created a more unstable environment than normal in which to potentially navigate middle school. Most obviously, many districts continued to use remote (or hybrid) instruction for large portions of 2020–2021 (IES, nd-a). Students experienced real barriers to remote learning such as insufficient access to school supplies, technology, the internet, and a quiet space at home. Even if students were in-person, students still experienced atypical schooling conditions. For example, they were subject to COVID-19 quarantine requirements and districts only had scattered policies in place to continue learning for these students (IES, nd-a). Student attendance fell (Chang et al., 2022), and perhaps because school routines were disrupted, students suffered increased feelings of isolation, depression, and anxiety (American Academy of Pediatrics, 2021). There are some small signs the pandemic positively impacted the learning environment for at least some students. For one, bullying decreased (Bacher-Hicks et al., 2022). On balance, however, evidence generally suggests the scattered nature of schooling may have made it more difficult for students and teachers to establish classroom practices, build a sense of community, and cultivate a positive and inclusive school climate—all things we know matter for student engagement and learning (Kutsyuruba et al., 2015; Lee, 2012; Wenglinsky, 2002).

The many pandemic-related changes to schooling outlined above likely affected at least some of these mechanisms through which schools seek to influence students' academic achievement. Many mechanisms likely broke down to some degree. For example, it was probably more difficult than normal for all students and teachers to foster and maintain personal relationships, regardless of school configuration and class size. At the same time, it is unlikely all mechanisms broke down entirely, perhaps especially in combined school settings. For example, there were likely already foundational personal relationships present among middle grade students in combined schools during the pandemic. In contrast, these relationships did not previously exist for students in standalone middle schools. Furthermore, whether schooling occurred remotely or in-person likely influenced to what degree these mechanisms broke down. Continuing the previous example, those in remote environments may have struggled to foster student-student and student-teacher relationships more so than their peers who were in-person. In summary, pandemic-era schooling conditions may have compromised the central idea that combined schools are predicated upon—that stable school environments which prioritize existing relationships can serve as a counterbalance to new academic difficulties and developmental challenges—but to an unknown degree.

Current Study

Pandemic-era schooling conditions created an atypical, though useful, context in which to explore the effectiveness of combined schools as a policy intervention. To our knowledge, this is the first study to examine how combined schools fared under these conditions. To find out whether the protective benefits historically offered by combined schools held up during the COVID-19 pandemic, we investigated math and reading achievement for youth in grades 6 and 7 who did and did not make structural middle school transitions. With data on roughly 20,000 students, this paper uses an event study design with propensity score matching to compare the achievement of students who attended combined schools (treatment) with those who underwent structural transitions (control). We examine the following research questions: (1) What was the effect of attending a combined school on students' math and reading achievement in spring 2021 (grade 6) and spring 2022 (grade 7), and how does this compare with the effect of attending this school structure in the pre-pandemic era? and (2) Did the effect of attending a combined school during the pandemic differ for sixth and seventh grade students learning predominately remotely versus predominately in-person in 2020–2021?

To preview our findings, we find that the advantage of attending a combined school diminished during the COVID-19 pandemic. Before the pandemic, sixth and seventh graders enrolled in combined schools had math and reading scores 0.12 to 0.16 standard deviation (SD) units higher than their peers in standalone middle schools. However, middle grade youth in combined schools in 2020–2021 and 2021–2022 had math and reading scores that were only 0.04 to 0.08 SDs higher than their transitional counterparts. School configuration mattered more for students' reading, but not their math, achievement if they were attending school in-person than if they were attending school remotely.

Our specific focus on atypical schooling conditions allows us to make two key contributions to the literature. First, we challenge the conventional wisdom that combined schools are necessarily effective at protecting student achievement and show this effectiveness is context dependent. Second, we illustrate that the combined school configuration itself can be insufficient for the academic benefits for students to be realized. Instead, our study suggests the combined school configuration likely functions as a catalyst that enables school-based relationships and other positive features of the learning environment that, in turn, provide educational benefits to students.

Materials and Methods

Data Sources

NWEA's Measures of Academic Progress (MAP) Growth K–8 math and reading assessments are computer adaptive, vertically scaled across grades, and are on a standard scale across states, which allows users to make comparisons over time and across jurisdictions (NWEA, 2019). MAP Growth is administered in the fall, winter, and spring of each school year. Schools and districts across the United States opt in to participating in MAP Growth assessments each school year. They also choose which test administrations to participate in and which of their students get assessed. According to NWEA, MAP Growth assessments have wide coverage nationally, including some 13 million students across 35,900 schools and 4,500 districts. Nevertheless, the opt-in model (combined with many students moving between schools) means many students in the MAP Growth database do not have consistent test observations that span all school years between kindergarten and grade 8.

Despite sometimes scattered student coverage, MAP Growth data have a major advantage relative to state assessments in that they span the pandemic era. There is no state assessment data from spring 2020 due to pandemic-related school closures and suspensions of state assessment systems. Meanwhile, spring 2021 state summative assessment data are presumed to suffer from low data quality due to inconsistencies in testing environments as well as in populations of students assessed. In contrast, MAP data include assessment scores in winter 2020, very limited assessment scores in spring 2020, and assessment scores of more typical quality in spring 2021.[1] For these reasons, MAP data have been used in multiple national analyses tracking achievement trends across the pandemic period (e.g., Kuhfeld & Lewis, 2022).

NWEA provided MAP Growth assessment data (RIT scores) spanning between the 2013–2014 school year and the 2021–2022 school year. NWEA also provided information on the schools and grades in which students were enrolled at the time of testing as well as student demographic information. We merged this MAP assessment data with the 2020–2021 Common Core of Data (CCD) to obtain information on the grades offered in each school, which allowed us to determine whether students made a structural transition in 2020–2021. From the CCD, we also obtained information on the school context, including the locale and poverty level as proxied by the proportion of students eligible for participation in the National School Lunch Program. We also merged in data from the COVID-19 Data Hub (CDH) dataset to measure learning modality in 2020–2021. Because it was common for schools to switch modalities (Kaufman & Diliberti, 2022) and because the frequency of modality data collection in the CDH varied by state, we used a simple dummy variable of whether schools were generally in-person or remote/hybrid for the majority of the 2020–2021 school year to create a consistent measure.

Sample Construction. We focused our analysis on students who made structural middle school transitions in grade 6 in 2020–2021 for two reasons: (1) it was the first time structural middle school transitions occurred after the COVID-19 pandemic first closed schools and (2) it allowed us to investigate how differences in instructional mode during the 2020–2021 school year might have differentially impacted students. This is also by far the most prevalent grade at which students nationally undergo a structural middle school transition. To retain focus on this transition, we linked students' MAP assessment data forwards and backwards in time, retaining students' spring assessment scores in grades 2 (2016–2017) through 7 (2021–2022). Grade 5 is a critical data year because it is the last school year before a theoretical transition. Our students were in grade 5 in 2019–2020, a time period that coincides with the initial school closures. Because we generally only had winter 2020 data available for these grade 5 students,[2] we use students' winter MAP scores in 2019–2020 but their spring scores in all other school years.[3] We exclude students for whom we did not have complete trend data, who were not observed in both math and reading, who did not make standard grade promotions, and who made structural middle school transitions in other grades (most often at grade 7).

Matching Strategy. Our analyses compare achievement for two groups of students: those who attended combined schools and those who made a structural middle school transition. We refer to these students as *treatment* and *control* students, respectively. We refer to combined schools as the treatment condition because they are an intentional policy intervention enacted because they are believed to provide better academic outcomes than their standalone counterparts. As the fundamental question of this paper is whether this policy intervention worked in a COVID-19 context, we believe it is most appropriate to use these students as the treatment condition. There are systematic differences in the types of students who tend to make a structural middle school transition and those who do not. This selection bias interferes with our ability to observe the effect of combined school enrollment on achievement because we are unable to determine whether differences in achievement are due to something about differences in the school configuration or to the systematic sorting of students across these school configurations. To address this bias, we assumed selection is driven by observable student and school demographic differences and used 1:1 nearest-neighbour propensity score matching to match students who attended combined schools with students who transitioned into a standalone middle school (Caliendo & Kopeinig, 2008).[4] Specifically, we predicted students' likelihood of being in a combined school conditional on various person- and school-level characteristics using the following logistic regression model:

$$T_i = \beta_0 + \beta_1 P_i + \beta_2 S_i + \varepsilon_i$$

where T_i is a dummy variable that equals 1 if student i attended a combined school during the study period and equals zero if the student did make a structural transition at grade 6. P_i is a vector of person-level characteristics (i.e., race and student achievement in math and reading in Grade 2). S_i is a vector of school-level characteristics that includes school locale (e.g., urban, suburban) and percentage of students eligible for free- and reduced-price lunch.[5] Notably, many more students attend standalone middle schools than attend combined schools, allowing for large numbers of theoretical matches.[6] However, we were unable to match some treated students due to missing demographic data.

Finally, we constructed a parallel, prepandemic analytic sample. This sample is comparable in all ways to the pandemic sample, with the exception that we did not need to use winter assessment scores in any data years due to missing spring scores. In this sample, data are shifted backward three school years, such that students enter the dataset in Grade 2 in 2013–2014 and the potential structural transition into grade 6 would occur in 2017–2018. This corresponds to the most recent cohort of students whose test scores in grades 2 through 7 were not impacted by the pandemic and is consistent with other studies that have used MAP Growth data to compare pandemic and prepandemic trends (see, e.g., Kuhfeld & Lewis, 2022).[7] We use both the prepandemic and pandemic analytic samples to address research question 1, and the pandemic sample only to answer research question 2 (see below for more details).

Our matched analytic pandemic sample includes 7,778 students and the prepandemic sample includes 12,506 students.[8] Although matching on observables significantly reduces our sample size, it makes our treatment and control groups much more similar at least on observables (see Table 1); that is, we consistently observe small standardized mean differences between treated and control students, below IES's standards for baseline equivalence (IES, nd-b).[9]

Table 1. Covariate Balance Table Before and After Propensity Scores Are Applied.

	Prepandemic analytic sample													
	Full sample							Matched sample						
	Control (*N* = 86,462)		Treatment (*N* = 17,400)			Sig?		Control (*N* = 6,253)		Treatment (*N* = 6,253)			Sig?	
Characteristic	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*
Math														
Grade 2 (S14)	193.2	(12.0)	190.6	(12.3)	0.21	***	***	190.4	(12.6)	190.4	(12.2)	0.00		
Grade 3 (S15)	205.9	(13.0)	204.4	(13.4)	0.11	***	***	203.2	(13.4)	203.5	(13.3)	−0.02		
Grade 4 (S16)	215.9	(13.9)	213.8	(14.0)	0.15	***	***	213.0	(14.4)	212.3	(14.2)	0.05	**	**
Grade 5 (S17)	225.2	(16.1)	222.7	(16.5)	0.15	***	***	222.1	(16.4)	221.2	(16.3)	0.06	**	**
Reading														
Grade 2 (S14)	190.5	(14.9)	187.3	(15.7)	0.21	***	***	187.2	(15.6)	187.2	(15.0)	0.00		
Grade 3 (S15)	201.7	(14.6)	198.9	(15.4)	0.18	***	***	198.7	(15.4)	198.7	(15.2)	0.00		
Grade 4 (S16)	209.6	(14.3)	207.2	(15.2)	0.16	***	***	206.6	(15.0)	206.2	(15.2)	0.02		
Grade 5 (S17)	215.6	(14.2)	213.4	(15.0)	0.15	***	***	212.8	(14.9)	212.3	(14.9)	0.03		
Demographics														
White	0.58	(0.49)	0.29	(0.45)	0.62	***	N/A	0.48	(0.50)	0.47	(0.50)	0.02		N/A
Black	0.10	(0.29)	0.18	(0.38)	−0.24	***	N/A	0.07	(0.26)	0.08	(0.27)	−0.03		N/A
Hispanic	0.14	(0.34)	0.40	(0.49)	−0.63	***	N/A	0.24	(0.43)	0.25	(0.43)	−0.01		N/A
Asian	0.04	(0.19)	0.05	(0.21)	−0.04	***	N/A	0.04	(0.21)	0.05	(0.21)	−0.01		N/A
Other	0.15	(0.35)	0.08	(0.28)	0.20	***	N/A	0.16	(0.36)	0.15	(0.36)	0.02		N/A
Urban	0.24	(0.42)	0.70	(0.46)	−1.06	***	N/A	0.27	(0.45)	0.31	(0.46)	−0.07	***	N/A
Suburban	0.43	(0.50)	0.12	(0.32)	0.74	***	N/A	0.27	(0.44)	0.26	(0.44)	0.02		N/A

(continued)

Table 1. (continued)

	Prepandemic analytic sample														
	Full sample							Matched sample							
	Control (*N* = 86,462)		Treatment (*N* = 17,400)			Sig?		Control (*N* = 6,253)		Treatment (*N* = 6,253)			Sig?	
Characteristic	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*
Town	0.12	(0.32)	0.02	(0.13)	0.41	***	N/A	0.03	(0.18)	0.03	(0.17)	0.02		N/A
Rural	0.22	(0.41)	0.16	(0.37)	0.15	***	N/A	0.43	(0.49)	0.41	(0.49)	0.04	*	N/A
Percent FRPL	0.47	(0.28)	0.55	(0.30)	−0.27	***	***	0.55	(0.29)	0.55	(0.30)	0.00		***

	Pandemic analytic sample														
	Full sample							Matched sample							
	Control (*N* = 75,186)		Treatment (*N* = 4,796)			Sig?		Control (*N* = 3,889)		Treatment (*N* = 3,889)			Sig?	
Characteristic	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*
Math														
Grade 2 (S17)	192.8	(12.7)	192.7	(12.3)	0.01		**	193.1	(12.5)	192.9	(11.7)	0.01		
Grade 3 (S18)	204.6	(13.0)	204.3	(12.7)	0.02		*	204.9	(12.7)	204.6	(12.0)	0.03		*
Grade 4 (S19)	214.6	(14.4)	214.0	(14.4)	0.04	**	**	214.8	(14.0)	214.3	(13.7)	0.03		*
Grade 5 (S20)	218.0	(14.7)	217.0	(14.6)	0.07	***	***	218.1	(14.4)	217.4	(13.9)	0.05	*	**
Reading														
Grade 2 (S17)	189.6	(15.6)	189.9	(15.2)	−0.02		*	190.4	(15.4)	190.3	(14.5)	0.01		*

(continued)

Table 1. (continued)

	Pandemic analytic sample													
	Full sample							Matched sample						
	Control (*N* = 75,186)		Treatment (*N* = 4,796)			Sig?		Control (*N* = 3,889)		Treatment (*N* = 3,889)			Sig?	
Characteristic	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*	*Mean*	*SD*	*Mean*	*SD*	*SMD*	*t-Test*	*KS Test*
Grade 3 (S18)	200.7	(15.0)	200.7	(14.8)	0.00			201.2	(14.6)	201.3	(14.1)	−0.01		
Grade 4 (S19)	207.9	(14.5)	207.7	(14.5)	0.01			208.3	(14.3)	208.4	(13.7)	−0.01		
Grade 5 (S20)	211.8	(14.0)	211.1	(14.2)	0.05	***	**	212.0	(13.8)	211.6	(13.4)	0.03		
Demographics														
White	0.53	(0.50)	0.51	(0.50)	0.04	**	N/A	0.51	(0.50)	0.51	(0.50)	0.00		N/A
Black	0.11	(0.32)	0.09	(0.28)	0.10	***	N/A	0.10	(0.30)	0.09	(0.29)	0.02		N/A
Hispanic	0.18	(0.38)	0.12	(0.32)	0.17	***	N/A	0.12	(0.33)	0.13	(0.34)	−0.02		N/A
Asian	0.04	(0.20)	0.02	(0.14)	0.11	***	N/A	0.02	(0.13)	0.02	(0.13)	−0.01		N/A
Other	0.14	(0.34)	0.26	(0.44)	−0.32	***	N/A	0.25	(0.43)	0.24	(0.43)	0.00		N/A
Urban	0.28	(0.45)	0.19	(0.39)	0.23	***	N/A	0.19	(0.39)	0.19	(0.39)	0.01		N/A
Suburban	0.43	(0.49)	0.33	(0.47)	0.21	***	N/A	0.32	(0.46)	0.32	(0.47)	−0.01		N/A
Town	0.09	(0.29)	0.06	(0.24)	0.13	***	N/A	0.07	(0.25)	0.06	(0.25)	0.02		N/A
Rural	0.19	(0.39)	0.42	(0.49)	−0.51	***	N/A	0.42	(0.49)	0.43	(0.49)	−0.01		N/A
Percent FRPL	0.51	(0.29)	0.45	(0.28)	0.22	***	***	0.44	(0.27)	0.45	(0.28)	−0.03		***

Note. SD = standard deviation, SMD = standardized mean difference. KS = Kolmogorov-Smirnov test. N/A = not applicable. Statistical significance calculated from t-tests and from Kolmogorov-Smironov tests on continuous variables only.

*p < .05, **p < .01, ***p < .001.

We perform several comparisons to assess selection bias in our prepandemic and pandemic analytic samples and to assess generalizability to national population of students. First, we confirm that the MAP-participating students that we excluded from our analytic samples (e.g., those who did not make standard grade promotions, those that made a structural transition at another grade level, those for whom complete trend data are unavailable) have similar student demographics and attended similar types of schools as those students who we retained in our analytic samples (see Table B1). We observe larger differences in the pandemic analytic sample. Compared with the full MAP Growth sample, students in our analytic sample are slightly more likely to be White, enrolled in rural schools, and tend to be enrolled in schools in which fewer students on average are eligible for free or reduced-price lunch. Second, we compare the achievement scores of students in our analytic sample to the national population of participating MAP students. We find our analytic sample is slightly higher achieving, particularly in the pandemic era (see Table B2). Third, we compare students in our analytic samples on observed student- and school-level demographics with the national student population (see Table B3). We find that relative to the national population of U.S. public school students in grade 6 in 2020–2021, our matched analytic samples have an over-representation of white students. Hispanic students are under-represented, especially in treatment schools. Furthermore, nationally, students in combined schools tend to be in urban areas, while the students in our matched analytic samples who attend combined schools disproportionately attend rural schools. We conclude from these checks that the students in our analytic sample are comparatively more advantaged than students nationally, particularly in the pandemic era. This suggests our analysis may overestimate the academic achievement of students in the pandemic era in both treatment and control schools. However, because this bias is present among both students in treatment and control schools, it is difficult to assess whether to be concerned that our treatment effects are biased upwards or downwards.

Plan of Analysis

To estimate the effect of combined school enrollment on student achievement in grades 6 and 7 we used an event study design. For RQ1, we estimated a series of linear regression models:

$$Y_{its} = \beta_0 + \beta_1 treatment_{its} + \sum_{\substack{g=2\\ g\neq 5}}^{7} \gamma\, grade_{its} + \sum_{\substack{g=2\\ g\neq 5}}^{7} \tau\, treatment_{its}\, x\, grade_{its} + \beta_2 race_{its} + \beta_3 locale_{its} + \beta_4 perfrl_{its} + \varepsilon_{its}$$

where Y_{its} is the assessment score (RIT score) of student i in school s at time t; $treatment_{its}$ is a time invariant dummy variable that equals 1 if student i attended a combined school throughout the study period and zero if the student instead made a structural middle school transition; $grade_{its}$ represents a series of dummy variables indicating the grade of student i at the time the test observation occurred; and $treatment_{its}$ $x\,grade_{its}$ is an interaction term. Meanwhile, $race_{its}$ is the race of student i while $locale_{its}$ represents the urbanicity of the school student i attended at time t and $frpl_{its}$ represents the share of students in the school eligible for free or reduced-price lunch in the school student i attended at time t. We estimated this model separately for math and reading achievement and separately for our prepandemic and pandemic samples. In all models, we used grade 5 as the reference time period. We clustered standard errors at the student level. The interaction terms (τ) capture differences in mean assessment scores between treated and control students at each grade level relative to Grade 5. We divided treatment effects by the corresponding subject x grade standard deviation from NWEA's, 2020 MAP Growth Norms to obtain an effect size (NWEA, 2020). We rely on guidance from Cohen (1969) and Kraft (2019) to guide interpretations of effect sizes.

To investigate RQ2 about differential effects based on the predominate learning modality used during the 2020–2021 school year, we added a time invariant covariate ($remote_{its}$) to our regression model that captures the learning modality of school that student i attended in the 2020–2021 school year. The variable $remote_{its}$ = 1 if the student attended a school that primarily used remote or hybrid instruction throughout the 2020–2021 school year and = zero if the school used primarily in-person instruction. We interact this variable with the $treatment_{its}$ $x\,grade_{its}$ terms. For ease of interpretation, we also convert our regression results to adjusted predictions (expected RIT scores) using Stata's *margins* command.

Results

Research Question 1

Using our event study models, we find evidence that students in treatment and control schools had achievement scores in the elementary grades (2–5) in both the prepandemic and pandemic eras that were not substantively (and generally not statistically significantly) different, especially in reading (see Table B4 and Figure 1). However, at grade 6 we begin to see evidence of a divergence in achievement scores between students who made a structural transition and students who did not. In the prepandemic era, we found that the students in our sample who remained in a combined school had MAP scores in grade 6 that were 2.35 points higher in math and 1.86 points higher in reading than students who underwent a structural middle school transition. By grade 7, this

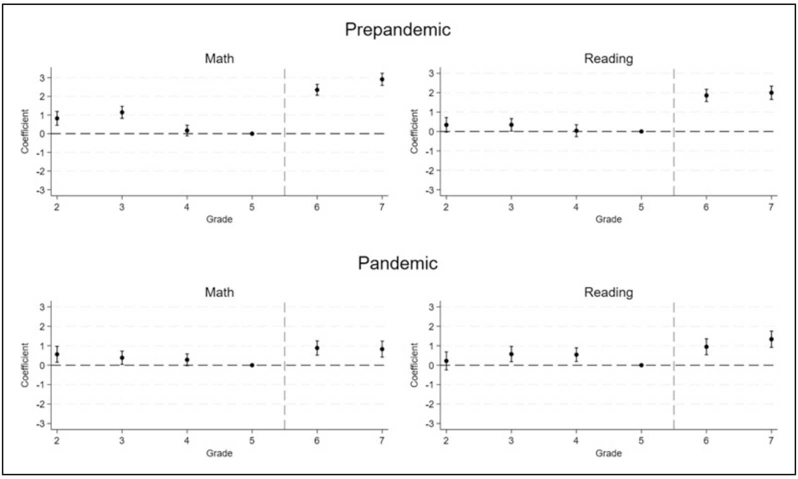

Figure 1. Event Study Coefficient Plot. *Note. N* = 12,506 in the prepandemic sample and 7,778 in the pandemic sample. Coefficients presented in this figure are the treatment x grade coefficients from main event study regression models (see Table B4). Bars represent 95 percent confidence intervals.

gap in achievement scores was 2.91 and 2.00 points in math and reading, respectively. Converting these differences to standardized units to assist with interpretation, we find students who attended combined schools in the prepandemic era had math and reading scores that were 0.12–0.16 standard deviation (SD) units higher in grades 6 and 7 than their transitional counterparts (see Table 2).

A different story emerged in our pandemic sample. While the advantage of combined school attendance did not disappear entirely in 2020–2021, it noticeably diminished. That is, we found that the achievement difference between students who remained in combined schools and those who made a structural transition was substantively smaller in both grades and subjects in the pandemic era than in the prepandemic era.[10] For example, students who remained in combined schools in 2020–2021 had math scores in grade 6 that were statistically higher than their transitional peers (0.88 points, or roughly 0.05 SD units). In grade 7, treated students continued to have math scores that were statistically higher (about 0.83 points or 0.04 SD units) than their peers who underwent a structural transition. However, the corresponding math score gaps in the prepandemic sample were 2.35 points (0.13 SD units) in grade 6 and 2.91 points (0.16 SD units) in grade 7. In summary, though we observed statistically significant differences in math and reading scores in the pandemic era between students in combined schools and those who made a structural

Table 2. Effect Sizes by Subject, Grade, and Student Subgroup.

	Prepandemic				Pandemic			
	Math		Reading		Math		Reading	
Student subgroup	Treatment Effect	Effect size[a]	Treatment Effect	Effect size[a]	Treatment Effect	Effect size[a]	Treatment Effect	Effect size[a]
Grade 6	2.35***	0.13	1.86***	0.12	0.88***	0.05	0.95***	0.06
Grade 7	2.91***	0.16	2.00***	0.12	0.83***	0.04	1.34***	0.08

*p < .05, **p < .01, ***p < .001.

[a]Effect size is treatment effect divided by the standard deviation (SD). We obtain SDs for a grade/subject in the spring term from NWEA's national norms: math, Grade 6 = 17.5; math, Grade 7 = 18.6; reading, Grade 6 = 16.0; reading, Grade 7 = 16.4 (NWEA, 2020).

transition, these differences were substantively smaller than the corresponding differences among our prepandemic sample.

Research Question 2

This research question asks whether the effect of attending a combined school on students' math and reading assessment scores differed for sixth and seventh grade students learning predominately remotely versus in-person. As a reminder, the remote indicator captures students' predominate learning modality during the 2020–2021 school year, when they were in grade 6. (Virtually all students, regardless of their learning modality in grade 6, were back in-person in grade 7.) Using our pandemic analytic sample, we found that the combination of learning modality and school grade configuration mattered differently for math versus reading (see Table B6).

Looking first at math, we observed consistently lower achievement in grades 6 and 7 for students who learned remotely for large portions of 2020–2021 relative to their in-person counterparts. This pattern held among both students who attended combined schools and those who made a structural transition. Figure 2—which plots regression-adjusted RIT scores by subject, grade, treatment status, and learning modality—clearly shows a divergence in achievement between students who learned remotely during 2020–2021 and

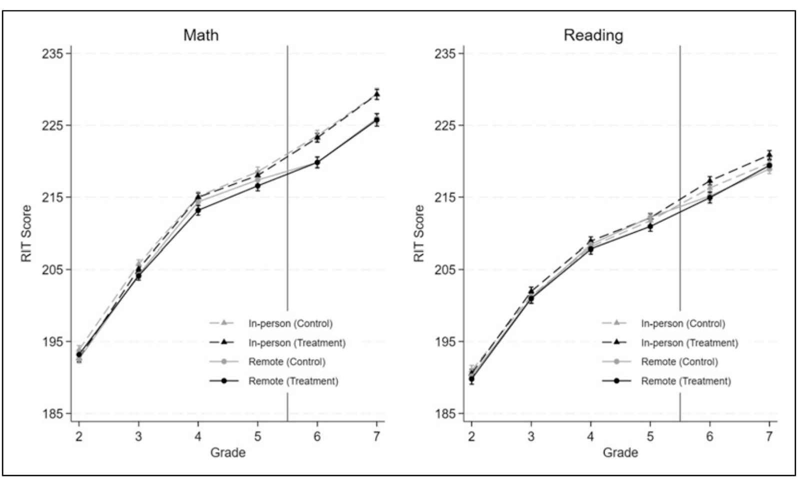

Figure 2. Mean RIT Score on MAP Assessments in the COVID-19 Pandemic Era by Subject, Grade, Treatment Status, and Learning Modality. *Note. N* = 6,758. Results are adjusted predictions (expected RIT scores) on MAP assessments by subject, grade, treatment status, and learning modality obtained from Stata's margins command after regression estimation.

those who learned in-person. However, we did not find evidence that the effect of attending a combined school on student achievement differed by whether students were learning predominantly remotely or in-person during the 2020–2021 school year. Said another way, conditional on learning modality, students in combined (treatment) schools had math assessment scores in grades 6 and 7 that were not statistically different from their counterparts that made a structural middle school transition. We interpret these findings as possible evidence that school grade configuration did not matter that much for middle grade students' math achievement during the pandemic era, perhaps because schools' choice of instructional mode mattered so much.

As shown in Figure 2, the patterns for students' reading assessment scores are more difficult to parse. Here, we did find some evidence that the combination of learning modality and school configuration influenced students' reading achievement. In Figure 3, we zoom in on students' regression-adjusted reading assessment scores in grades 6 and 7 by treatment status and learning modality to help explain these patterns. Students who attended combined (treatment) schools in-person had statistically higher reading assessment scores in Grade 6 than their peers who also attended combined schools but did so remotely. These sixth graders in in-person combined (treatment) schools also marginally outperformed their in-person counterparts who made a structural transition (control). In contrast, sixth graders who attended combined schools remotely had reading assessment scores that were

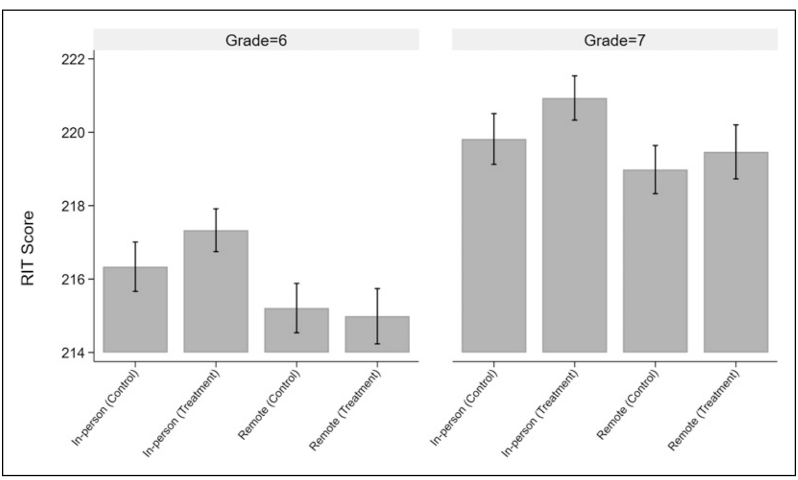

Figure 3. Mean RIT Score on MAP Reading Assessments During the COVID-19 Pandemic Era by Grade, Treatment Status, and Learning Modality. *Note. N* = 6,758. Results are adjusted predictions (expected RIT scores) on MAP reading assessments by grade, treatment status, and learning modality obtained from Stata's margins command after regression estimation.

statistically no different than their peers who attended school remotely after making a structural transition. This continued into grade 7. That is, in grade 7, students who had learned remotely the previous year in a combined school had reading assessment scores that were still no higher than their counterparts who had made a remote structural middle school transition. But seventh graders who had remained in-person in combined schools continued to slightly outperform their in-person transitional counterparts. In fact, by grade 7, students who remained in-person in combined schools during the 2020 had the highest reading assessment scores of any group. In summary, school configuration mattered more for students' reading achievement if they were attending school in-person than if they were attending school remotely. However, even the observed reading score differences by learning modality and school configuration were substantively small.

Discussion and Implications

Many prepandemic studies have found that combined schools can protect students from the academic declines associated with making a structural middle school transition. Using an event study design, we too found evidence that students who remained in combined elementary and middle schools in the prepandemic era had higher achievement scores in both math and reading in the middle grades than students who made a structural middle school transition. Our estimated prepandemic treatment effects are around 0.12 SD units in reading (equivalent to 1.86 to 2.00 points) and 0.13 to 0.16 SDs in math (2.35–2.91 points, respectively). This is roughly on par with estimates from other studies conducted with prepandemic samples (Atteberry et al., 2022; Dhuey, 2013; Schwerdt & West, 2013). Taken altogether, this suggests combined schools were doing an adequate job protecting students' academic outcomes before the COVID-19 pandemic. However, our analyses newly find that the protective benefits historically offered by combined schools diminished during the pandemic. Even though we did observe positive treatment effects in the pandemic era, these effects are consistently smaller than the corresponding effects in the prepandemic era. Notably, we observed equivalent or even larger treatment effects in grade 7—when students were back learning in-person and schooling conditions were more stable—than in grade 6. Previous studies (Dhuey, 2013; Hong et al., 2018; Rockoff & Lockwood, 2010) have found mixed evidence about how long achievement differences persist beyond the transition grade.

Perhaps surprisingly, we did not observe strong evidence of differential treatment effects by the predominate instructional mode students received used in 2020–2021. We might have expected in-person combined schools to offer the greatest protective benefits in the pandemic era since this group is closest to normal schooling conditions. While we did see suggestive evidence that this was true at least in reading, the differences are not as stark as we might have predicted. We hypothesize this is because there were widespread disruptions to schooling in 2020–2021 and, to a lesser extent, in 2021–2022 (IES, nd-a), including in schools that were predominately in-person. These disruptions may have destabilized schools' normal operating conditions, even in in-person combined schools. For math, we found that learning modality mattered more than school configuration for student achievement during the pandemic era. That is, regardless of whether students underwent a structural transition or remained in a combined school, students who attended school remotely for large portions of 2020–2021 had lower math achievement in grades 6 and 7 than their in-person peers. This echoes findings from other studies which have found remote instruction to be a key driver of math achievement coming out of the pandemic (Goldhaber et al., 2022).

Our central finding that the protective benefits historically offered by combined schools diminished during the pandemic era has three key implications for education policy and practice. First, our study has implications for understanding the protective benefits of combined schools during future schooling disruptions. We do *not* interpret our findings as evidence that policymakers should abandon the combined school configuration in favor of a standalone model; our analysis with the prepandemic sample confirms this is an effective educational intervention under normal schooling conditions. That said, this study has implications for policy failures we can expect to occur when disruptions to schooling inevitably happen in the future. The COVID-19 pandemic is an extreme example of a prolonged, widespread disruption to schooling, which makes for an interesting case study. But disruptions to schooling happen all the time. In the last decade, some 13 million students were negatively impacted by school closures due to things like extreme weather, natural disasters, violence, teacher strikes, (Jahan et al., 2022). Unplanned school closures are only expected to become more common in the future (Interagency Network for Education in Emergencies, 2020), especially as climate change exacerbates extreme weather events (Petek, 2022). Future disruptions to schooling will inevitably affect students in combined school environments. Our findings suggest policymakers cannot assume the protective benefits typically offered by this school configuration hold during these periods of disruption. Said another way, the stable school environment that combined schools seek to promote will not happen accidentally during these disruptions. Instead, this environment will need to be deliberately created and nurtured. When atypical schooling conditions do inevitably occur in the future, our findings suggest that educators and policymakers should work to mimic stability and continuity in their school environment as best as possible. This might mean engaging in activities that prioritize maintaining existing student-student and student-teacher relationships, like retaining existing class groupings or increasing the practice of "looping," where students stay with their same teacher for multiple grades.

Relatedly, our findings have a second implication for typical schooling conditions, regardless of school structure. Although we find that the protective benefits offered by combined schools largely evaporated during the pandemic, not all the mechanisms that likely create these protective benefits completely broke down. This suggests it is not the school configuration itself offering benefits, but the influence of one (or more) specific mechanisms typically disproportionately present in combined schools. There are multiple mechanisms that potentially make combined schools different than standalone middle schools (e.g., continuous teacher and peer relationships, established norms and expectations). As Atteberry et al., 2022 too recommends, future research should seek to isolate exactly *which* mechanism (or mechanisms) in combined schools is causing the academic benefits. If we are able identify which mechanism is actually doing the work, we can prioritize preserving the integrity of this mechanism during future periods of school disruptions. More importantly, we can work to port this mechanism over to the more prevalent grade 6–8 and 7–9 school environments so other students can reap the academic benefits. For example, if combined schools offer academic benefits primarily via continuous teacher-student relationships, this might suggest middle schools should increase the practice of "looping"—not just during periods of disruption but during typical schooling conditions as well.

And finally, there are the implications for our understanding of how the pandemic impacted students' learning and the associated implications for students' academic recovery. Documenting how the pandemic impacted students' academic outcomes is critical to inform decisionmaking about how to help students recover. State and local policymakers need to determine where to invest resources (e.g., tutoring programs) and need to identify which students in what grade levels and in what subjects are suffering the most. For example, recent evidence of COVID-19 impacts on achievement have helped researchers and policymakers reach agreement that students' middle school math achievement must be a clear focus of recovery efforts. We cannot assume the combined school environment protected students against learning lost during the COVID-19 pandemic. The middle school students who were outperforming their peers before the pandemic may not be now. Thus, combined schools might need to newly be a focus area for pandemic recovery investments.

Limitations

Readers should keep in mind several limitations. Though our sample is large and diverse, it includes only students in schools who opted to participate in MAP Growth. Thus, our results may not necessarily be representative of the entire population of public school students. Second, emerging research shows that public school enrollments shifted during the COVID-19 pandemic; families moved to different public schools or left the public education entirely (Dee, 2023; Dee & Murphy, 2021). These pandemic-related enrollment shifts raise two related concerns. For one, the population of students in our pandemic sample is slightly different from our prepandemic sample, at least measured by observable characteristics (see Table 1). We are unable to tease out whether the different effects we observe in our prepandemic and pandemic samples are simply due to differences in sample compositions. Furthermore, differential enrollment shifts might have contributed to differential attrition in our pandemic sample in ways that might bias our results. In sum, we cannot separate out the impact of pandemic-related enrollment shifts on achievement from our effect of interest. Finally, although we control for school urbanicity, we did not control for geographic differences in how the pandemic impacted schooling (or for the pandemic directly).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research reported here was sponsored by the Institute of Education Sciences, U.S. Department of Education, through grant R305U200006.

ORCID iD

Melissa K. Diliberti https://orcid.org/0009-0002-6035-3097

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Many districts were still offering remote (or hybrid) instruction for large portions of the 2020–2021 school year (IES, nd-a). A minority of districts administered spring 2021 assessments under these conditions, creating concerns about MAP data comparability from assessments administered in-person versus those administered remotely. Kuhfeld et al. (2020) did a comparison analysis and concluded remote testing data was trustworthy for middle school grades.
2. Spring 2020 assessment scores are available for a very small share of students (about 5 percent of the students who are assessed in a typical spring). Setting aside

the concern that conducting this analysis with this 5 percent of students would lead to results that would be representative of the broader population of students, reliance on these spring 2020 assessment scores would not leave us with a large enough sample size to conduct this analysis since so few students—both nationally and in the MAP Growth database—attend combined schools. Thus, although limited spring 2020 assessment scores are available, they are unsuitable for analysis.

3. We are cautious about using students' winter 2020 scores for two reasons: (1) fewer students participate in MAP Growth testing in the winter, reducing our possible sample size, and (2) we were concerned students' winter 2020 scores were systematically lower than their theoretical spring 2020 scores if these students had been tested. To account for these concerns, as a robustness check, we imputed students' spring 2020 assessment scores. We modeled our imputation approach after Schweig et al., (2021) which found that regression-based replacement strategies using MAP Growth data can generally produce scores that are highly correlated with students' true scores. We imputed students' spring 2020 assessment scores as a function of their on-subject and off-subject prior achievement on spring assessments, school average prior achievement, and student demographics (see Schweig et al., 2021). We replicated our main analysis using students' imputed spring 2020 assessment scores in grade 5 and confirm our main results (see Table B5).
4. We matched treated and control students on their propensity scores using four methods: 1) one-to one nearest neighbour without replacement, 2) one-to-one nearest neighbour with replacement, 3) k nearest neighbours, and 4) kernel matching. All nearest neighbour matching methods performed similarly well, notably better than the kernel matching method, presumably due to the large number of potential matches. After reviewing results from multiple matching methods, we used results from the one-to-one nearest neighbour match without replacement because this method has the most intuitive, straightforward interpretation. Our results are not sensitive to which matching method we choose.
5. We use school-level data on student eligibility for free or reduced-price meals as a proxy for poverty status because we do not have student-level data.
6. As shown in Atteberry et al. (2022), K–8 schools have expanded over the last decade. In our pandemic sample, only 6 percent of students were enrolled in combined schools, and in our prepandemic sample only 17 percent were enrolled in combined schools. We believe the decline in the share of our sample enrolled in K–8 schools between the prepandemic and pandemic eras has to do with changes in the schools participating in MAP Growth assessments. However, we do not know how the pandemic might have influenced schools' or districts' decisions about whether to participate in MAP Growth assessments.
7. We chose this cohort as our prepandemic cohort to align with MAP Growth studies examining learning during the COVID-19 pandemic relative to the prepandemic era. We could have chosen another prepandemic comparison cohort. We do not

believe our choice of reference cohort matters for several (related) reasons. First, our cohort of interest is the pandemic sample; this prepandemic cohort is only intended to be a reference point—constructed using the same methods as our pandemic sample—so that there is appropriate context to interpret our pandemic results. Second, a large body of literature confirms that before the COVID-19 pandemic, students in K–8 schools had higher achievement in the middle grades than their transitional counterparts. We observe this exact same pattern and find effect sizes of a comparable magnitude to many other prepandemic studies using a variety of data years. Thus, we do not believe at our choice of which prepandemic cohort to use meaningfully affects our results.

8. Differences in analytic sample sizes are likely due to changes in the schools and districts participating in MAP Growth over time.
9. We present all results using the matched analytic samples which adjust for selection bias on observables. However, because our propensity score matching technique significantly reduces our sample size, as a robustness check, we replicate our main analysis using our full analytic samples. Using the larger analytic sample where we do not match students on propensity scores, we observe the same pattern in math—smaller treatment effects in grades 6 and 7 in the pandemic era relative to the prepandemic era. We observe the same general pattern in reading too, but the differences between the prepandemic and pandemic eras are larger in math. Results available upon request.
10. We formally test this using a triple difference-in-difference set up. Specifically, we incorporate an indicator variable into our regression models that captures whether student i belongs to our prepandemic or pandemic sample. We interact this indicator variable with our $treatment_{its}\ x\, grade_{its}$ term. The coefficients on these triple interaction terms capture whether there are differential impacts of treatment in grades 6 and 7 in the prepandemic versus pandemic eras. Using this set up, we confirm there are differential treatment effects in the prepandemic versus the pandemic era at both grades 6 and 7 in math ($p < 0.001$) and reading (grade 6, $p < 0.001$; grade 7, $p < 0.01$).

References

Alspaugh, J. W. (1998). Achievement loss associated with the transition to middle school and high school. *The Journal of Educational Research*, *92*(1), 20–25. https://doi.org/10.1080/00220679809597572

Alspaugh, J. W. (2000). The effect of transition grade to high school, gender, and grade level upon dropout rates. *American Secondary Education*, *29*(1), 2–9.

American Academy of Pediatrics. (2021). *AAP, AACAP, CHA declare national emergency in children's mental health. [Press release]*. https://www.aap.org/en/advocacy/child-and-adolescent-healthy-mental-development/aap-aacap-cha-declaration-of-a-national-emergency-in-child-and-adolescent-mental-health/

American Enterprise Institute. (2021). *Return to learn tracker. [Data set]*. https://www.returntolearntracker.net/instructional_status/

Anderman, E. M., & Kimweli, D. M. (1997). Victimization and safety in schools serving early adolescents. *The Journal of Early Adolescence*, *17*(4), 408–438. https://doi.org/10.1177/0272431697017004004

Atteberry, A., Wedow, R., Cook, N. J., & McEachin, A. (2022). A replication of a quasi-experimental approach to estimating middle school structural transition effects on student learning trajectories. *Educational Policy*, *36*(7), 1612–1651. https://doi.org/10.1177/089590482098779

Bacher-Hicks, A., Goodman, J., Green, J. G., & Holt, M. K. (2022). The COVID-19 pandemic disrupted both school bullying and cyberbullying. *The American Economic Review*, *4*(3), 353–370. https://doi.org/10.1257/aeri.20210456

Barber, B. K., & Olsen, J. A. (2004). Assessing the transitions to middle and high school. *Journal of Adolescent Research*, *19*(1), 3–30. https://doi.org/10.1177/0743558403258113

Byrnes, V., & Ruby, A. (2007). Comparing achievement between K–8 and middle schools: A large-scale empirical study. *American Journal of Education*, *114*(1), 101–135. https://doi.org/10.1086/520693

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Carnegie Council on Adolescent Development. (1989). *Turning points: Preparing American youth for the 21st century*. Carnegie Corporation of New York.

Chang, H., Balfanz, R., & Byrnes, V. (2022). *Pandemic causes alarming increase in chronic absence and reveals need for better data*. Attendance Works Blog. https://www.attendanceworks.org/pandemic-causes-alarming-increase-in-chronic-absence-and-reveals-need-for-better-data/

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences* (1st ed.). Academic Press.

Dee, T. S. (2023). *Where the kids went: Nonpublic schooling and demographic change during the pandemic exodus from public schools*. The Urban Institute. https://www.urban.org/research/publication/where-kids-went-nonpublic-schooling-and-demographic-change-during-pandemic

Dee, T. S., & Murphy, M. (2021). Patterns in the pandemic decline of public school enrollment. *Educational Researcher*, *50*(8), 566–569. https://doi.org/10.3102/0013189X211034481

Dhuey, E. (2013). Middle school or junior high? How grade-level configurations affect academic achievement. *Canadian Journal of Economics/Revue canadienne d'économique*, *46*(2), 469–496. https://doi.org/10.1111/caje.12020

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *48*(2), 90–101. https://doi.org/10.1037//0003-066x.48.2.90

Eccles, J. S., Lord, S., & Midgley, C. (1991). What are we doing to adolescents? The impact of educational contexts on early adolescents. *American Journal of Education*, *99*(1), 52–542. https://doi.org/10.1086/443996

Eccles, J. S., & Midgley, C. (1988). Stage/environment fit: Developmentally appropriate classrooms for early adolescents. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education*. Academic Press.

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence*, *21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x and https://psycnet.apa.org/doi/10.1111/j.1532-7795.2010.00725.x

Goldhaber, D., Kane, T., McEachin, A., Morton, E., Patterson, T., & Staiger, D. (2022). The consequences of remote and hyrid instruction during the pandemic. *NBER Working Paper 30010*.

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology*, *17*(3), 300–312. https://psycnet.apa.org/doi/10.1037/0012-1649.17.3.300

Hong, K., Zimmer, R., & Engberg, J. (2018). How does grade configuration impact student achievement in elementary and middle grades? *Journal of Urban Economics*, *105*(1), 1–19. https://doi.org/10.1016/j.jue.2018.02.002

Institute of Education Sciences. (nd-b). What works clearinghouse standards brief. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_baseline_080715.pdf

Institute of Education Sciences. (nd-a). Highlights from the 2021 NAEP monthly school Survey. https://ies.ed.gov/schoolsurvey/mss-report/

Interagency Network for Education in Emergencies. (2020). 20 Years of INEE: Achievements and challenges in education in emergencies. https://inee.org/resources/20-years-of-inee

Jahan, F. A., Zviedrite, N., Gao, H., Ahmed, F., & Uzicanin, A. (2022). Causes, characteristics, and patterns of prolonged unplanned school closures prior to the COVID-19 pandemic—United States, 2011–2019. *PLoS One*, *17*(7), Article e0272088. https://doi.org/10.1371/journal.pone.0272088

Juvonen, J. (2007). Reforming middle schools: Focus on continuity, social connectedness, and engagement. *Educational Psychologist*, *42*(4), 197–208. https://doi.org/10.1080/00461520701621046 and https://psycnet.apa.org/doi/10.1080/00461520701621046

Juvonen, J., Le, V., Kaganoff, T., Augustine, C. H., & Constant, L. (2004). *Focus on the wonder years: Challenges facing the American middle school*. RAND Corporation.

Kaufman, J., & Diliberti, M. K. (2022). *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*. *RAND*. https://www.rand.org/pubs/research_reports/RRA168-6.html.

Kraft, M. (2019). *Interpreting effect sizes of education interventions*. Annenberg EdWorkingPaper No. 19-10.

Kuhfeld, M., & Lewis, K. (2022). *Student achievement in 2021–2022: Cause for hope and continued urgency. NWEA*. https://www.nwea.org/uploads/2022/07/Student-Achievement-in-2021-22-Cause-for-hope-and-concern.researchbrief-1.pdf.

Kuhfeld, M., Lewis, K., Meyer, P., & Tarasawa, B. (2020). *Comparability analysis of remote and in-person MAP Growth testing in fall 2020*. NWEA. https://www.nwea.org/uploads/2020/03/Technical-brief-Comparability-analysis-of-remote-and-inperson-MAP-Growth-testing-in-fall-2020-NOV2020.pdf

Kutsyuruba, B., Klinger, D. A., & Hussain, A. (2015). Relationships among school climate, school safety, and student achievement and well-being: A review of the literature. *The Review of Education*, *3*(2), 103–135. https://doi.org/10.1002/rev3.3043

Lee, J. S. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. *International Journal of Education Research*, *53*(1), 330–340. https://doi.org/10.3350/cmh.2012.18.3.330

Malone, M., Cornell, D., & Shukla, K. (2017). Association of grade configuration with school climate for 7th and 8th grade students. *School Psychology Quarterly*, *32*(3), 350–366. https://doi.org/10.1037/spq0000174 and https://psycnet.apa.org/doi/10.1037/spq0000174

Mergendoller, J. R. (1993). Introduction: The Role of research in the reform of middle Grades education. *The Elementary School Journal*, *93*(5), 443–446. https://doi.org/10.1086/461733

NWEA. (2019). MAP® Growth™ technical report.

NWEA. (2020). 2020 NWEA MAP Growth normative data overview.

Offenberg, R. (2001). The efficacy of Philadelphia's K-to-8 schools compared to middle Grades schools. *Middle School Journal*, *32*(4), 23–29. https://doi.org/10.1080/00940771.2001.11495283

Petek, G. (2022). *Climate change impacts across California. K–12 education*. State of California Legislative Analyst's Office. https://lao.ca.gov/Publications/Report/4586

Rockoff, J. E., & Lockwood, B. B. (2010). Stuck in the middle: Impacts of grade configuration in public schools. *Journal of Public Economics*, *94*(11–12), 1051–1061. https://doi.org/10.1016/j.jpubeco.2010.06.017

Schwartz, A. E., Stiefel, L., & Rothbart, M. W. (2016). Do top dogs rule in middle school? Evidence on bullying, safety, and belonging. *American Educational Research Journal*, *53*(5), 1450–1484. https://doi.org/10.3102/0002831216657177

Schwartz, A. E., Stiefel, L., Rubenstein, R., & Zabel, J. (2011). The path not taken: How does school organization affect eighth-grade achievement? *Educational Evaluation and Policy Analysis*, *33*(3), 293–317. https://doi.org/10.3102/0162373711407062

Schweig, J., McEachin, A., Kuhfeld, M., Mariano, L. T., & Diliberti, M. K. (2021). *Adapting course placement processes in response to COVID-19 disruptions:*

Guidance for schools and districts. *RAND*. https://www.rand.org/pubs/research_reports/RRA1037-1.html.

Schwerdt, G., & West, M. R. (2013). The impact of alternative grade configurations on student outcomes through middle and high school. *Journal of Public Economics*, *97*(1), 308–326. https://doi.org/10.1016/j.jpubeco.2012.10.002

Sebastian, C., Burnett, S., & Blakemore, S. J. (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences*, *12*(11), 441–446. https://doi.org/10.1016/j.tics.2008.07.008

Seidman, E., Allen, L., Aber, J. L., Mitchell, C., & Feinman, J. (1994). The impact of school transitions in early adolescence on the self-system and perceived social context of poor urban youth. *Child Development*, *65*(2 Spec No), 507–522.

Weiss, C. C., & Kipnes, L. (2006). Reexamining middle school effects: A comparison of middle grades students in middle schools and K–8 schools. *American Journal of Education*, *112*(2), 239–272. https://doi.org/10.1086/498996 and https://psycnet.apa.org/doi/10.1086/498996

Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, *10*(12), 12. https://doi.org/10.14507/epaa.v10n12.2002

Author Biographies

Melissa Kay Diliberti, PhD, is an Associate Policy Researcher at the RAND Corporation. Her research interests include domestic and international education policy, and her recent work has focused on using educator surveys to examine the impact of the COVID-19 pandemic on the U.S. education system.

Jonathan D. Schweig, PhD, is a Senior Social and Behavioral Scientist at the RAND Corporation. His research focuses on education policy and the measurement of instructional practices, classroom and school climate, and social and emotional competencies.

Louis T. Mariano, PhD, is a Senior Statistician at the RAND Corporation whose education research has focused on evaluation of the efficacy of education programs, policies and reforms; experimental and quasi-experimental design methodology; and statistical applications to mental measurement, including student assessments.

Megan Kuhfeld, PhD in Education, is a Research Manager at NWEA. Her research seeks to understand students' academic and social-emotional learning trajectories and the school and neighborhood influences that promote optimal growth.

Andrew Mceachin, PhD in Education Policy, is the Senior Research Director of Policy Research at ETS. His research focuses on understanding the causes and consequences of educational inequities.