

# Ready or Not?

## A New Way to Measure Elementary and Middle School Quality

By Jing Liu, Ph.D., Seth Gershenson, Ph.D.  
and  
Max Anthenelli



Foreword by  
David Griffith  
and  
Amber M. Northern

February 2025

THOMAS B.  
**FORDHAM**  
INSTITUTE  
ADVANCING EDUCATIONAL EXCELLENCE

## About Fordham

The Thomas B. Fordham Institute promotes educational excellence for every child in America via quality research, analysis, and commentary, as well as advocacy and exemplary charter school authorizing in Ohio. It is affiliated with the Thomas B. Fordham Foundation, and this publication is a joint project of the Foundation and the Institute. For further information, please visit our website at [fordhaminstitute.org](https://fordhaminstitute.org). The Institute is neither connected with nor sponsored by Fordham University.

## Suggested Citation

Jing Liu, Ph.D. Seth Gershenson, Ph.D. and Max Anthenelli. Ready or Not? A New Way to Measure Elementary and Middle School Quality. Washington D.C.: Thomas B. Fordham Institute (February 2025). <https://fordhaminstitute.org/national/research/ready-or-not-new-way-measure-elementary-and-middle-school-quality>

## Acknowledgments

This report was made possible by our sister organization, the Thomas B. Fordham Foundation, as well as by numerous individuals whose efforts are reflected in the final product. First among those equals are the authors, Jing Liu, Seth Gershenson, and Max Anthenelli, whose persistence and professionalism was essential to the project's success. We also thank Dan Goldhaber and Jeff Denning for their clear and thoughtful feedback on the research methods. And here at Fordham, we thank Chester E. Finn, Jr. and Michael J. Petrilli for providing feedback on the draft, Stephanie Distler for managing report production and design, and Victoria McDougald for overseeing media dissemination.

## Contents

Foreword.....	3
Introduction .....	11
Background .....	13
Data .....	14
Methodology.....	15
Findings.....	16
Implications.....	22
Appendix A: Methodology.....	23
Appendix B: Additional Tables.....	24
Endnotes .....	28

## Foreword

By David Griffith and Amber M. Northern

Everyone knows that standardized tests paint an incomplete picture of schools' impacts on students. Yet, the search for alternative indicators of school quality has been a disappointment.

That's because coming up with thoughtful alternatives to tests is easier said than done. After all, any truly informative indicator of school quality must satisfy at least five criteria.

*First, it must be **valid**.* In other words, it must capture something that policymakers, parents, or "we the people" care about.

*Second, it must be **reliable**.* That is, it must be reasonably accurate and consistent from one year to the next or when averaged over multiple years.

*Third, it must be **timely**.* For example, it makes little sense to report elementary schools' effects on college completion, since that information would be useless to parents and policymakers by the time it was available.

*Fourth, it must be **fair**.* For example, if the point is to gauge school performance, an indicator shouldn't systematically disadvantage schools with lots of initially low-performing students (though of course, it's also important to measure and report students' absolute performance).

*Finally, it must be **trustworthy**.* For example, it may be unwise to hold high schools accountable for their own graduation rates, as ESSA requires, since doing so gives schools an incentive to lower their graduation standards (and their grading standards).

Finding an alternative indicator that checks all of these boxes is inherently challenging. And in practice, bureaucratic inertia, risk aversion, and the difficulty of collecting and effectively communicating new data to stakeholders can make it hard for policymakers to think outside the box.

Still, there is at least one underutilized and potentially powerful data point that nearly every public school in the country already collects—a student's grade point average (GPA). Might it also be used to evaluate school quality?

Perhaps the most obvious objection to this idea is that (like holding high schools accountable for their own graduation rates) holding schools accountable for their own students' GPAs incentivizes grade inflation. But what if we instead rewarded schools for improving students' grades at the *next* institution they attend? For example, what if an *elementary* school's accountability rating were partly based on its students' GPAs in *sixth* grade? Similarly, what if a *middle* school's accountability rating were partly based on the grades that its graduates earned in *ninth* grade?

A measure that is based on students' subsequent GPAs has at least two compelling features.

First, because it captures students' performance across *all* subjects—not just reading and math—it provides parents and policymakers with valuable information about underappreciated dimensions of schools' performance.

Second, by holding schools accountable for students' success at their *next* school, it disincentivizes teaching to the test, socially promoting students, and other behaviors that yield short-lived or illusory gains.

Of course, like standardized test scores, the grades students earn reflect their lives outside of school, in addition to whatever grading standards teachers and schools choose to establish. But what if we held schools accountable for their *effects* on students' "subsequent" grades, much as traditional test-based value-added measures hold them accountable for their effects on achievement?

In the past decade, rigorous research has shed new light on the effects that individual schools have on non-test-based outcomes,<sup>[1]</sup> of which "subsequent GPA" is perhaps the most obviously connected to the day-to-day work of educators and has perhaps the most robust record of predicting students' success.<sup>[2]</sup> All else equal, students with higher GPAs do better,<sup>[3]</sup> which means it matters if some schools enhance (or degrade) the skills that help them earn good grades at their next institution.

Yet, to our knowledge, there has been no real discussion of how schools might be rewarded or penalized for the value they add to students' grade point averages. So, to better understand the potential of "GPA value-added" as an indicator of school quality, we asked two of the country's most prolific education scholars, the University of Maryland's Jing Liu (who was ably seconded by his research assistant, Max Anthenelli) and American University's Seth Gershenson, to conduct parallel analyses of nearly a decade of student-level data from Maryland and North Carolina, both of which collect detailed information on students' individual course grades.

Because of the limitations of that data, the North Carolina analysis does not examine elementary schools' GPA-value-added (whereas the Maryland analysis includes both elementary and middle schools). But overall, the results from the two states were encouraging and exhibit similarly desirable properties in other jurisdictions, should they choose to explore it.

### **So, how does it work?**

To isolate a middle school's contribution to students' 9th grade GPAs, the proposed measure controls for a broad range of variables including but not limited to individual students' 8th grade GPAs, 8th grade reading and math scores, and socio-demographic backgrounds, as well as the average 8th grade GPA of each individual middle school.<sup>[4]</sup> It then limits subsequent comparisons to students attending the same high school.

In practice, the inclusion of those school-average variables and the subsequent same-high school restriction effectively accounts for the unique approaches that schools take to grading, thus ensuring that middle schools that send their students to high schools with tougher grading standards aren't unfairly handicapped. But of course, even within a high school, not all courses are created equal. So, to avoid penalizing middle schools for putting more students on accelerated

tracks, the proposed indicator only assesses a school's effect on students who enroll in Algebra I in 9th grade, meaning it excludes the roughly 40 percent of students who enroll in more or less advanced courses.<sup>15</sup>

Intuitively, the resulting model relies on comparisons between "observably similar" students who attended *different* middle schools but the *same* high school (and were thus subject to the same grading standards). What's more, it averages across lots of teachers and even more student-teacher relationships, thus ensuring that a particular school's score won't be unfairly high or low because Ms. Johnson doesn't give homework or Mr. Smith doesn't like little Johnny's attitude.

Principals and teachers who are familiar with test-based growth measures might best understand the new indicator as "GPA-based growth," or the academic progress that a middle school's students make as measured by their high school GPAs (or that an elementary school's students make as measured by their middle school GPAs).

For parents and other guardians, an indicator of students' "high school readiness" (or "middle school readiness") likely makes more sense.

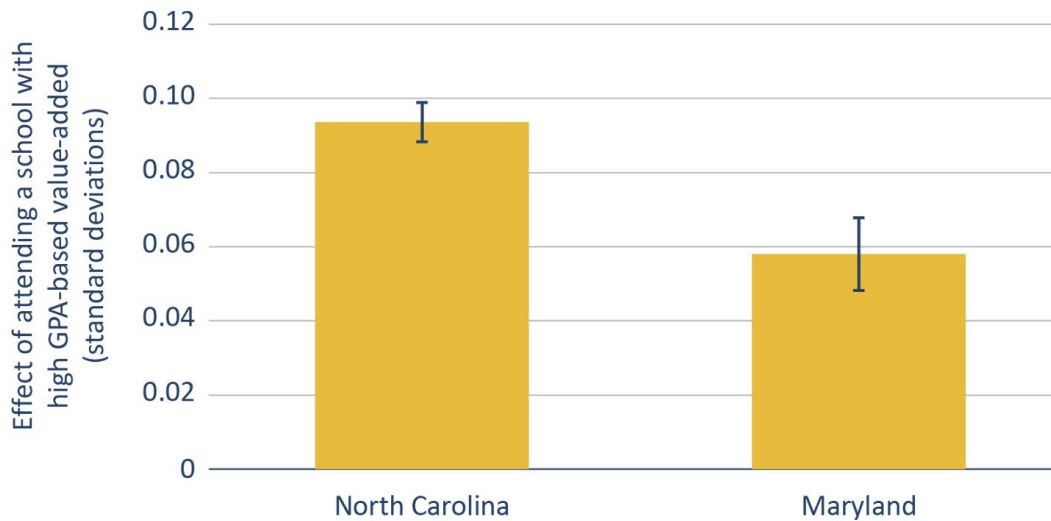
Now let's look at how well the proposed indicator satisfies the five criteria outlined above.

### Is it valid?

By definition, non-trivial effects on subsequent GPA are important. And in fact, the results presented in Finding 1 suggest that ***both elementary and middle schools have sizable effects on the grades that students earn at their next schools.*** For example, attending a middle school with strong "GPA growth" was associated with a 0.1 standard deviation increase in a North Carolina student's ninth grade GPA and a 0.06 standard deviation increase in a Maryland student's ninth grade GPA (Figure ES-1) – or about 0.1 grade points.

As every high school teacher knows, how a student fares in 9th grade typically forecasts his or her future success. So, it's no surprise that prior research finds that 9th grade GPA predicts 11th grade GPA<sup>16</sup> and that high school GPA in general is more predictive of postsecondary outcomes than test scores, at least for students who go to college.<sup>17</sup> Still, as discussed in Finding 3, ***even schools that boost short-run outcomes don't necessarily boost long-run outcomes such as high school graduation and college going,*** perhaps because graduation is a lousy metric, perhaps because both graduation and college going are inherently difficult for elementary and middle schools to change, or perhaps because our estimates are too imprecise to detect whatever effects do exist.

**Figure ES-1: Attending a middle school with high GPA-based value-added has a sizable effect on a student's 9th grade GPA.**

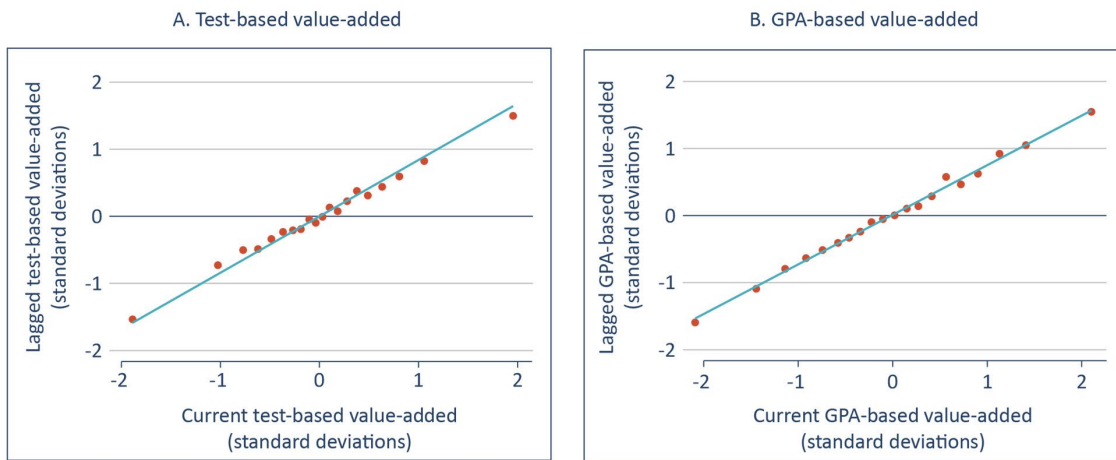


Notes: This figure shows the estimated effect of attending a middle school with a GPA value-added score that is one standard deviation above the average middle school's score on the average student's 9th grade GPA.

**Is it reliable?**

Overall, the results suggest GPA-based growth is about as reliable as test-based growth. For example, the “intertemporal correlation coefficient,” which is a measure of year-to-year stability, is 0.86 for test-based value-added in North Carolina middle schools and 0.74 for GPA-based value-added (Figure ES-2). In Maryland, the analogous estimates show that GPA-based value-added is actually more stable than its test-based counterpart.

**Figure ES-2: Like test-based growth, GPA-based growth is fairly stable over time.**



Notes: This figure shows the intertemporal correlation of value-added to 9th grade math achievement and value-added to 9th grade GPA for North Carolina middle schools. A stronger correlation indicates greater year-to-year stability, which is one indication of reliability.

### Is it timely?

As noted, we propose that elementary schools be rated based on their effects on students' *sixth* grade GPAs and that middle schools be rated based on students' *ninth* grade GPAs,<sup>[8]</sup> meaning that information on middle and high school readiness would be about a year "out of date" by the time it reached parents and policymakers.

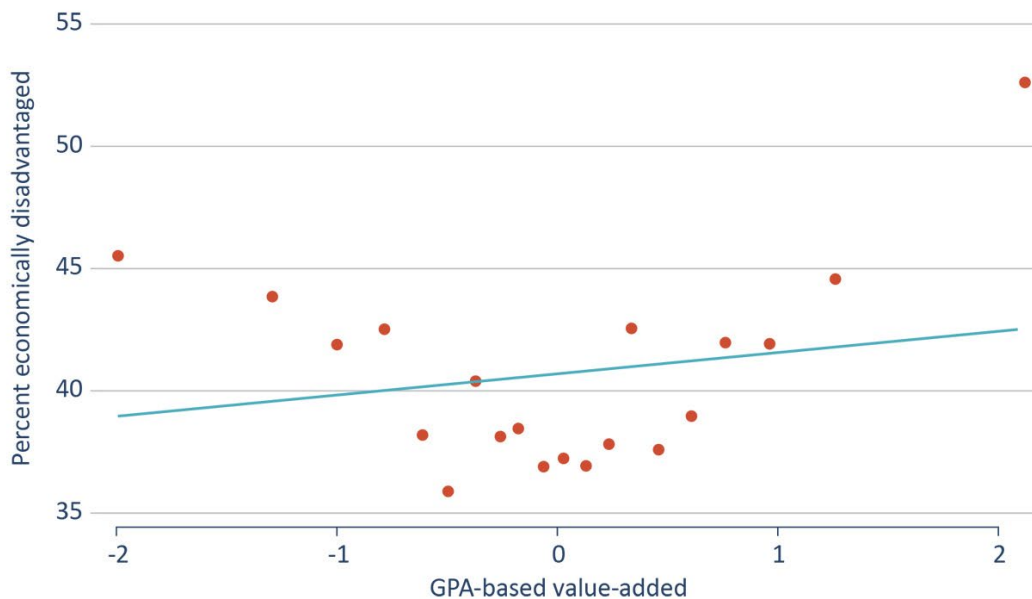
In our view, that's timely enough. For example, a parent choosing between middle schools whose students made similar progress in English language arts and math would likely be wise to send their child to the school with a higher "high school readiness" score (even if it was based on the *previous* year's graduates). Similarly, a policymaker weighing the pros and cons of intervening in a school with subpar test-based growth would probably be better informed if he or she had access to slightly out-of-date information on that school's GPA-based growth.

### Is it fair?

As indicated, both the elementary measure and the two middle school measures are essentially uncorrelated with observable demographic characteristics. For example, there is no statistically significant relationship between a Maryland middle school's GPA-based growth and the percentage of students who are economically disadvantaged (Figure ES-3).

Based on this result, our sense is that GPA-based growth gives teachers who work in challenging environments a fair shake.

**Figure ES-3: Schools' socio-demographic composition doesn't predict their GPA-based growth.**



Notes: This figure shows the correlation between value-added to 9th grade GPA and the percentage of students who are economically disadvantaged for Maryland middle schools.



**Is it trustworthy?**

Although it’s impossible to be sure until the measure has been tried, our sense is that manipulating an elementary or middle school’s GPA-based growth would be difficult. For example, the typical high school student gets letter grades in 6-8 subjects, each of which is taught by a different teacher. Moreover, even if every teacher in a high school lowered his or her grading standards, the effect on its feeder middle schools’ GPA-based growth would be minimal because the measure would still be relying on comparisons between students from different feeder middle schools.

Still, teachers and principals must be made aware of these facts, or they may be tempted to act on their misunderstanding. For example, high school teachers might try to “protect” their middle school colleagues by inflating students’ grades. So, before piloting or otherwise experimenting with the measure, we strongly recommend that states explain to educators the impossibility of artificially boosting schools’ GPA-based growth (and the detrimental consequences of attempting to do so). In our experience, most educators take seriously their role in preparing students for the next grade level—and this indicator doesn’t make them choose between that noble goal and fairness to their colleagues – but again, it’s up to education leaders to communicate that.

In sum, GPA-based growth mostly satisfies our five criteria for sensible indicators of school quality.

**How does “GPA-based growth” compare to other measures?**

Per Table 1, our sense is that GPA-based growth compares favorably to many existing measures of school quality. For example, any measure that controls for prior achievement and other student characteristics is fairer to educators than raw achievement or chronic absenteeism rates (though yes, students should also be held to a common standard). Moreover, both chronic absenteeism rates and “attendance value-added”<sup>[9]</sup> are potentially vulnerable to manipulation – for example, if schools don’t properly account for partial-day absences – though our sense is that technological advances now make tracking attendance more secure.

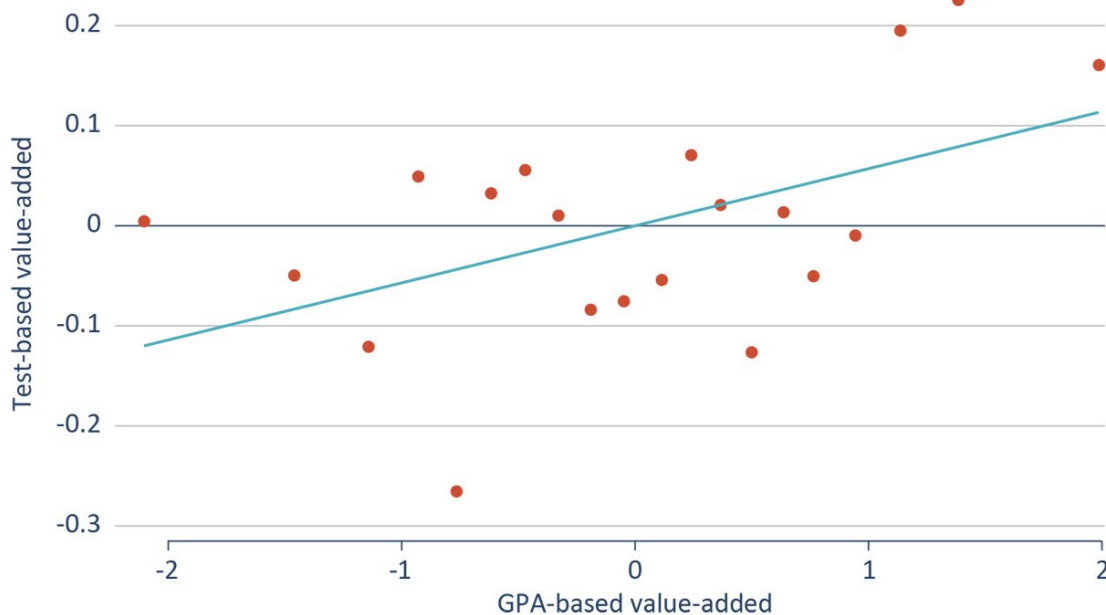
Table 1: Comparing indicators of elementary and middle school quality					
	Test-based indicators		Non-test-based indicators		
	Achievement/ Proficiency	Test-based growth	Chronic absenteeism	Attendance- based growth	GPA-based growth
<b>Valid</b>	✓	✓	✓	✓	✓
<b>Reliable</b>	✓	✓	✓	✓	✓
<b>Timely</b>	✓	✓	✓	✓	✓
<b>Fair</b>	X	✓	X	✓	✓
<b>Trustworthy</b>	✓	✓	?	?	✓

Notes: This table summarizes the pros and cons of various indicators of school quality according to Fordham authors. For more information on “attendance-based growth” (a.k.a., “attendance value-added”), see Jing Liu’s 2022 report: Imperfect Attendance: Toward a fairer measure of student absenteeism.

Still, when push comes to shove, incorporating GPA-based growth isn't really about how well it compares to other measures (most of which are worth keeping). It's about how much a K-8 rating system that relies mostly on reading and math scores and some derivative of attendance is likely to miss the mark if the goal is a rigorous understanding of how much a particular school contributes to student success.

That brings us to the report's last finding, which is that ***schools' effects on subsequent grades are weakly correlated with their effects on subsequent test scores***. Among other things, this means there are many schools with high test-based growth and low GPA-based growth and many schools with low test-based growth and high GPA-based growth (Figure ES-4).

**Figure ES-4: GPA-based growth is weakly correlated with test-based growth.**



Notes: This figure shows the correlation between value-added to 6th grade GPA and value-added to 6th grade math achievement for Maryland elementary schools.

In short, both the data and common sense suggest that a system that relies heavily on test-based value-added isn't capturing everything that we care about, even if the things it does capture are essential to students' success.

### What are the pitfalls?

Based on what we've learned thus far, we see three sticking points.

First, because the underlying model relies on comparisons between high school peers who attend different middle schools, the proposed measure won't work for places where all or nearly all high school students attend the same middle school—something that occurs more often in rural areas – though because most schools enroll at least a few students from schools that aren't feeder schools, the measure should work in many places with a single feeder school.<sup>[10]</sup>

Second, communicating how the measure is calculated is no easy task. So, it may be advisable for states to start by providing this information as supplemental data, rather than including it in their rating systems (assuming they have school ratings).

Finally, the measure is potentially incompatible with a strict reading of ESSA<sup>(11)</sup> (though as always, there's wiggle room). So, absent unforeseen changes to the federal landscape, states that have developed parallel rating systems or dashboards may be best served by considering how an appropriately vetted indicator of "GPA growth" or "High School Readiness" might be incorporated into *those* systems after it has been piloted.

To be clear, we are not advocating for any version of the proposed measure to *replace* test-based growth. However, we do believe that it has potential as a supplement.

So, let's experiment with this new measure. Because at the end of the day, it sends a clear message to schools that one of their core missions is to help their graduates succeed in their next step – not just in reading and math, but in *all* subjects – and not just on tests, but on the stuff that tests struggle to capture.

In short, it gives educators whose contributions are sometimes shortchanged by bubble sheets and multiple-choice questions an officially sanctioned reason to do something that everyone should want them to do.

Teach to the best of their ability.

## Introduction

Among the most important developments in education policy has been the rise of test-based growth measures as a supplement to traditional measures of student achievement. In recent years, these “value-added” measures have been used to award bonuses to highly effective teachers and schools, identify teachers and schools in need of remediation, and provide information about teacher and school quality to parents, policymakers, and other stakeholders.

Yet despite the apparent utility—and widespread use—of value-added techniques, it’s no secret that there’s a growing backlash against standardized testing. Related critiques, all of which have a dash of validity, include that “teaching to the test” effectively narrows the curriculum, that testing takes too much class time, that unavoidable idiosyncrasies arise in a one-off test that purports to measure a year’s worth of learning,<sup>[12]</sup> and (perhaps most important) that standardized tests inevitably paint an incomplete picture of students’ well-being and growth, as well as the overall value and effectiveness of the schools they attend.

For these and other reasons, the Every Student Succeeds Act (the 2015 update to No Child Left Behind) required states to adopt at least one “alternative” indicator of school quality, in addition to achievement, growth, progress towards English Language proficiency (for English Language Learners), and graduation (for high schools). By requiring an alternative measure, the architects of the law hoped that states would innovate and develop new indicators that paint a more well-rounded picture of school quality.

In theory, this approach made sense. Yet in practice, almost every state adopted an attendance-related metric as its “fifth indicator.” While attendance has some appealing features, such as its objective and easy-to-record nature, making a school’s attendance or chronic absenteeism rate the only non-test-based indicator of school performance is questionable for at least three reasons: First, insofar as the goal is gauging school performance, a focus on attendance levels has the same problems as a focus on academic achievement—chief among them that it is strongly correlated to school demographics. Second, even if states adopted some sort of growth measure for attendance that controlled for demographics (which has yet to happen),<sup>[13]</sup> a school’s influence on student attendance seems to be more muted than its influence on other academic outcomes.<sup>[14]</sup> Finally, and perhaps most importantly, attendance as such is only tangentially related to the core academic and socio-emotional skills that comprise the central mission of public schools (as any teacher with a punctual but otherwise disengaged student is all too aware).

The purpose of this report is to propose consideration of a broader and potentially more powerful alternative: grades.

As every parent knows, letter grades usually reflect a combination of students’ academic proficiency and socio-emotional or “non-cognitive” skills that are valued in a productive citizenry, such as participation, teamwork, timeliness, and attention to detail. Indeed, some early studies found that a student’s grade point average (GPA) in high school was more predictive of postsecondary and labor market success than his or her test scores—though the

evidence on this point is mixed.<sup>[15]</sup> Yet, despite the evidence that GPA captures something important, we know little about how schools contribute to it – or, to be more precise, how they affect students’ capacity to earn better grades in an educational setting with a particular grading standard.<sup>[16]</sup>

Because grading standards vary, it is challenging to investigate a school’s effect on its students’ “capacity” to earn good grades while they are still enrolled. And of course, it makes no sense to hold schools accountable for the grades they assign. However, by waiting to see how students fare at their next institution, where they and observably similar peers from other schools are held to a common grading standard, it is possible to solve both of these problems.

Accordingly, this study breaks new ground by developing measures of school quality that capture schools’ impacts on students’ subsequent grades – that is, their grades at the next institution in which they enroll. More specifically, we use longitudinal administrative data from two states, Maryland and North Carolina, to estimate elementary and middle schools’ effects on students’ 6th and 9th grade GPAs. In other words, we estimate schools’ “GPA value-added” in much the same way that prior work has estimated schools’ test-score or attendance value-added.<sup>[17]</sup> To accomplish this, we compare the GPAs of otherwise similar students who graduated from different elementary (or middle) schools but went to the same middle (or high) school, thus circumventing the concern that different schools have different grading standards.

Because our ultimate goal is to gauge the potential utility of our measures in real-world settings, we address the following research questions:

- 1. How much do elementary and middle schools affect students’ grades in sixth and ninth grade? In other words, how much do they vary in their “GPA value-added”?***
- 2. How much does an elementary or middle school’s GPA value-added vary over time?***
- 3. How strongly correlated are GPA value-added and test-score value-added?***
- 4. To what extent does GPA value-added predict other outcomes of interest, such as academic achievement, repeating 9th grade, high school graduation, and post-secondary plans or enrollment?***

Collectively, the answers to these questions suggest that calculating and reporting schools’ “GPA value-added” could provide parents and policymakers with valuable information about heretofore underappreciated dimensions of schools’ performance.

## Background

Objective and evidence-based measures of school quality are of paramount importance to the functioning of our education system. Among other things, they facilitate the spread of effective practices, the identification of schools in need of improvement, and informed decision-making on the part of parents, including—for a growing number—their choice of schools for their children. Yet, decades of education research have demonstrated that it is critical to distinguish between the quality of schools and the performance of their students, which is plausibly linked to numerous factors that are largely beyond the control of frontline educators. Hence, the rise of measures that seek to quantify schools’ “value-added” by taking students’ demographics and baseline performance into account.

Value-added models have their origins in the identification of individual teachers’ contributions to student learning, as measured by test scores.<sup>[18]</sup> However, they were eventually adapted to measure teachers’ effects on students’ non-test score outcomes as well as schools’ effects on both test and non-test student outcomes.<sup>[19]</sup> The common theme in all of these models is to adjust as completely as possible for pre-existing student characteristics and incoming achievement, or the things that are outside of schools’ and teachers’ control that may influence both school or classroom assignments and subsequent achievement. Once these adjustments are made, the average growth of students in a particular classroom or school (i.e., the teacher’s or school’s value added) can be compared to the average growth of students in other classrooms and schools.

The evidence is now quite clear that both teachers and schools vary considerably in their value-added to test- and non-test-based outcomes. For example, in addition to varying in their effects on student reading and math scores, schools vary in their effects on student attendance and social-emotional learning,<sup>[20]</sup> and GPA itself has been included in an index of “noncognitive skills” that teachers are known to influence.<sup>[21]</sup> Other research also suggests that the skills and behaviors that contribute to GPA are malleable. For example, student-facing “growth-mindset” and “sense-of-purpose” interventions have been shown to increase the GPAs of students at risk of failing to complete high school.<sup>[22]</sup> In short, it stands to reason that schools vary in their effects on subsequent GPAs, which likely reflect an amalgam of content mastery, study skills, effort, and other intangibles.

Many of the non-cognitive skills that employers and taxpayers value are difficult to measure directly.<sup>[23]</sup> But they are inherent in the grades that students earn, albeit to varying degrees.<sup>[24]</sup> And, unlike test scores, letter grades are almost always awarded in *all* subjects by the start of middle school. Consequently, GPA predicts a number of long-run outcomes. For example, ninth grade GPA predicts subsequent test-based achievement and eleventh grade GPA, which is important in college admission decisions.<sup>[25]</sup> Similarly, high-school GPA predicts college graduation. Indeed, it may be more predictive than students’ performance on college entrance exams.<sup>[26]</sup>

## Data

To enhance the generalizability of our findings, we conduct parallel analyses of similar state-level longitudinal data from both Maryland and North Carolina, thus eliminating the concern that the findings are attributable to the unique characteristics of a particular education system or data set.

Importantly, both data systems contain transcripts for each high school student, which we use to construct students' GPAs in their first year of high school (9th grade). Maryland also contains middle-school and elementary school transcripts, which we use to construct analogous measures for lower grades.

Although Local Education Agencies (LEAs) sometimes calculate GPA differently, for the purposes of this project we use a standard scale. That means that A, B, C, D, and F correspond to 4, 3, 2, 1, and 0, respectively, and pluses or minuses add or subtract  $\frac{1}{2}$  of a grade point (with the exception of an A+, which is still a 4). Because most LEAs add .5 grade points for honors courses and 1 grade point to AP and IB courses, we also construct a weighted GPA measure that follows this rule irrespective of LEA policy (though in practice, taking this step has little effect on the results).

Ultimately, our analyses of Maryland and North Carolina complement one another in at least two ways: First, Maryland is one of the most diverse and urbanized states in the U.S., with a student population that is now 33% White, 33% Black, 22% Hispanic/Latino, and 7% Asian. In contrast, much of North Carolina is mostly suburban or rural, and its K-12 student population is 48% White, 25% Black, 18% Hispanic/Latino, and 3% Asian. Second, because all K-12 data are linked to administrative postsecondary data in the Maryland Longitudinal Data System (MLDS), our analysis of the long-run effects of the measure includes college enrollment in Maryland. Because we cannot make this connection in North Carolina, our analysis of long-term effects there relies on survey data that include a student's plan to enroll in college (or not) upon high school graduation.

## Methodology

Our measures of “GPA-value added” are similar to the school and teacher test-score value added measures that are now common in K-12 educational research and policy.<sup>[27]</sup> The basic idea is to isolate an individual school’s contribution to its students’ subsequent outcomes by adjusting for the fact that different schools educate different types of students, which we accomplish using measures of students’ and schools’ prior academic performance and socioeconomic status. Because grading standards vary, we also make one further adjustment that is unique to GPA value-added: we adjust for the middle (high) school that the graduates of an elementary (middle) school attend. For example, we compare the ninth grade GPAs of students who attended different middle schools but the same high school, thus eliminating the concern that some middle schools are being held to a higher standard.

To provide some context for our results, in addition to estimating schools’ effects on subsequent GPA across all subjects, we estimate their effects on subsequent math scores.<sup>[28]</sup> After computing each school’s value-added to both subsequent GPAs and subsequent math test scores, we relate both of these value-added measures to a host of short- and long-run outcomes including end-of-course test scores, GPAs, 9th grade repetition, high school graduation, college intent, and college enrollment (see Appendix A for a detailed description of these techniques).



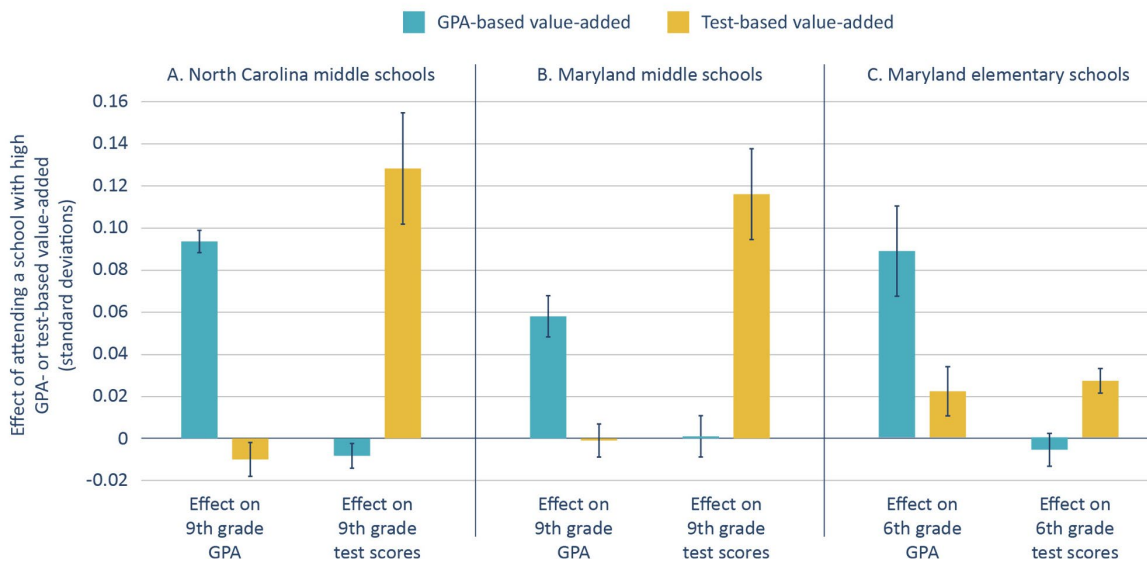
## Findings

### Finding 1: Both elementary and middle schools have sizable effects on the grades that students earn at their next schools.

The headline result of our study is that both elementary and middle schools have sizable effects on students’ subsequent GPAs (Figure 1). In North Carolina, we estimate that a one standard deviation (SD) increase in a middle school’s GPA value-added increases students’ ninth grade GPAs by almost 10 percent of a standard deviation (Panel A). In Maryland, we estimate that a one SD increase in a middle school’s GPA value-added increases students’ ninth grade GPA by almost 6 percent of a SD (Panel B), while a one SD increase in an elementary school’s GPA value added increases students’ sixth grade GPAs by 9 percent of a SD (Panel C).

Importantly, these effects are apparent even when the middle school’s test-score value-added is accounted for in the model.<sup>[29]</sup> However, per the figures, a middle school’s GPA value-added has a negligible effect on students’ ninth grade math scores. Similarly, a middle school’s test-based value-added has a significant and positive effect on students’ 9th grade math achievement, but a negligible effect on students’ 9th grade GPAs. Collectively, these results suggest that test-based and GPA-based value-added capture different dimensions of a school’s effectiveness – which isn’t too surprising, given that the latter captures a fundamentally different sort of performance across a far broader range of circumstances and subjects.

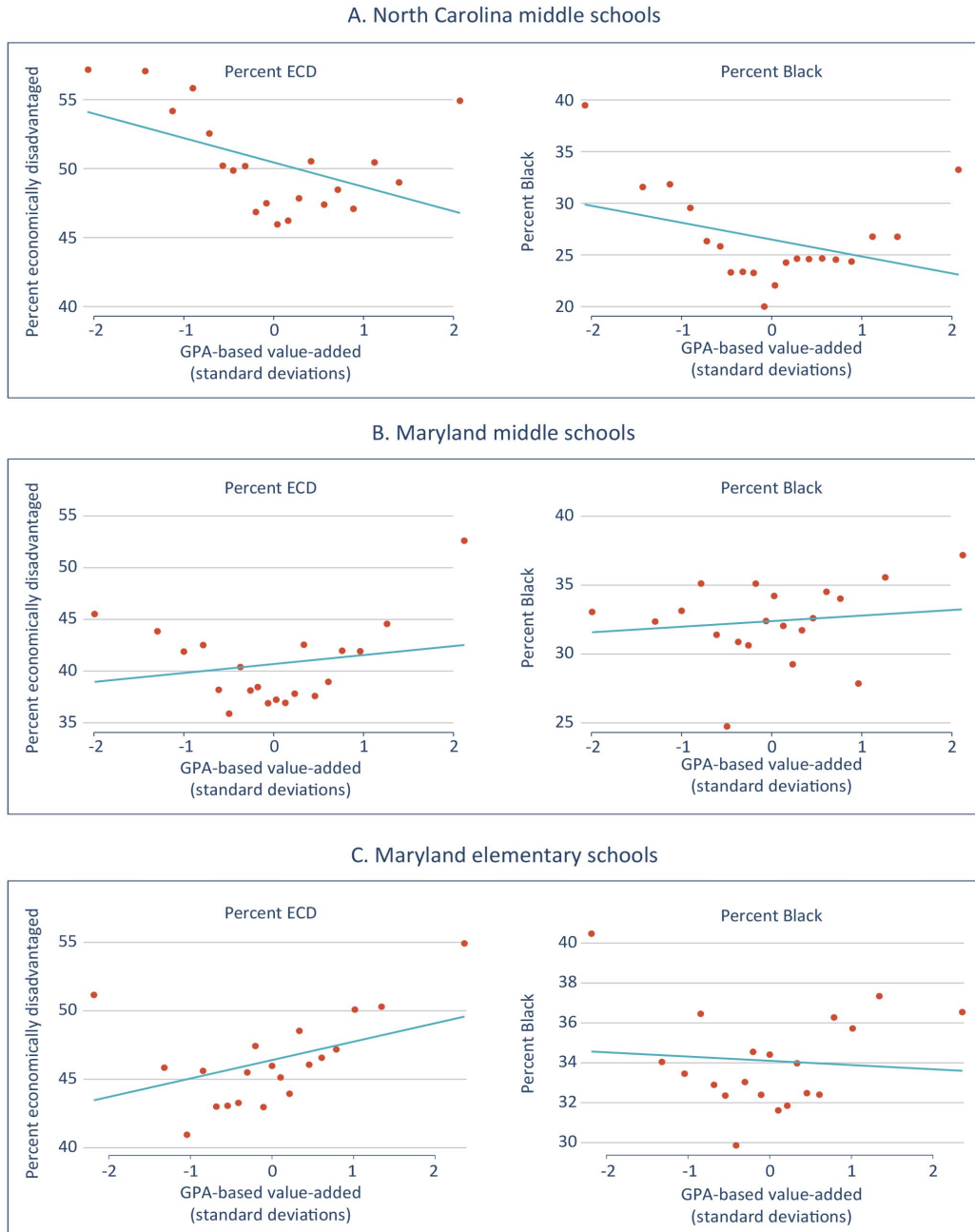
**Figure 1: Middle schools with high GPA value-added boost students' 9th grade GPAs, while those with high test-based value-added boost students' 9th grade test scores.**



Notes: This figure shows the estimated effect of attending an elementary or middle school with test- or GPA-based value-added that is one standard deviation above the average school’s on 6th or 9th grade math achievement and on the GPA of the average North Carolina or Maryland student.

Per Figure 2, there is no significant relationship between the percentage of a school’s students who are economically disadvantaged or Black and its GPA value-added in any of the samples analyzed in this study. Given that the underlying model controls for these demographic variables, this “null” finding is unsurprising. However, it is nevertheless reassuring insofar as it provides further evidence that the value-added measures capture actual school effects and not the fact that some schools serve more or less advantaged students.

**Figure 2: Schools’ socio-demographic composition does not predict their GPA value-added.**



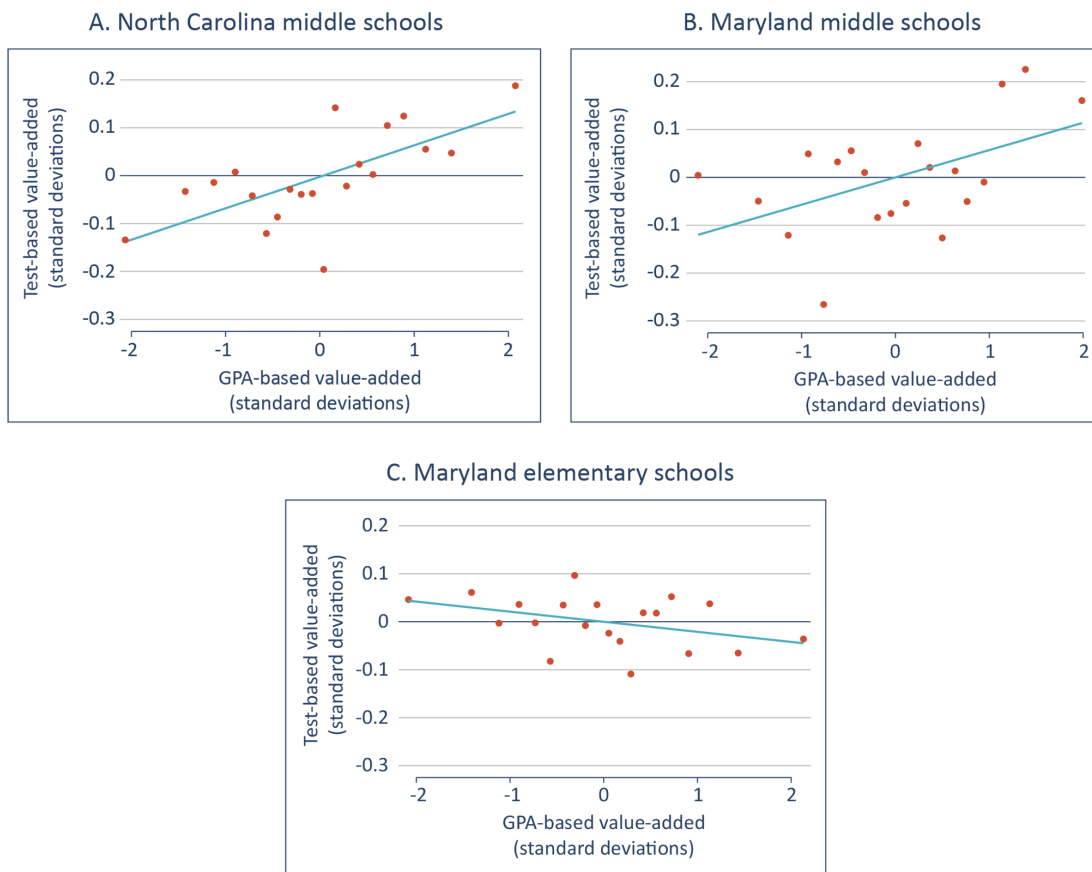
Notes: This figure shows the correlation between an elementary or middle schools’ GPA value-added and the percentage of students who are economically disadvantaged or Black in North Carolina and Maryland.

Finally, there is some evidence (not shown) that the effects of attending a school that improves students' grades are concentrated among students with lower baseline achievement. Specifically, attending a North Carolina middle school with high GPA value-added boosts the 9th grade GPA of students in the lowest tertile of 7th grade math achievement by 0.13 standard deviations (or more than twice as much as it boosts the 9th grade GPA of students in the top tertile of 7th grade math achievement). However, we don't see this pattern in the Maryland data.

**Finding 2: Schools' effects on subsequent grades are weakly correlated with their effects on subsequent test scores.**

Consistent with prior research on school<sup>[30]</sup> and teacher quality,<sup>[31]</sup> our estimates suggest that GPA value-added identifies different subsets of elementary and middle schools as high- and low-performing than test-based value-added does. In fact, we find almost no relationship between the two measures (Figure 3). For example, the correlation between middle schools' GPA- and test-based value-added in North Carolina is 0.07 (Panel A). And in Maryland, the equivalent figure is 0.06 for middle schools (Panel B) and just 0.01 for elementary schools (Panel C).

**Figure 3: GPA value-added is weakly correlated with test-based value-added.**



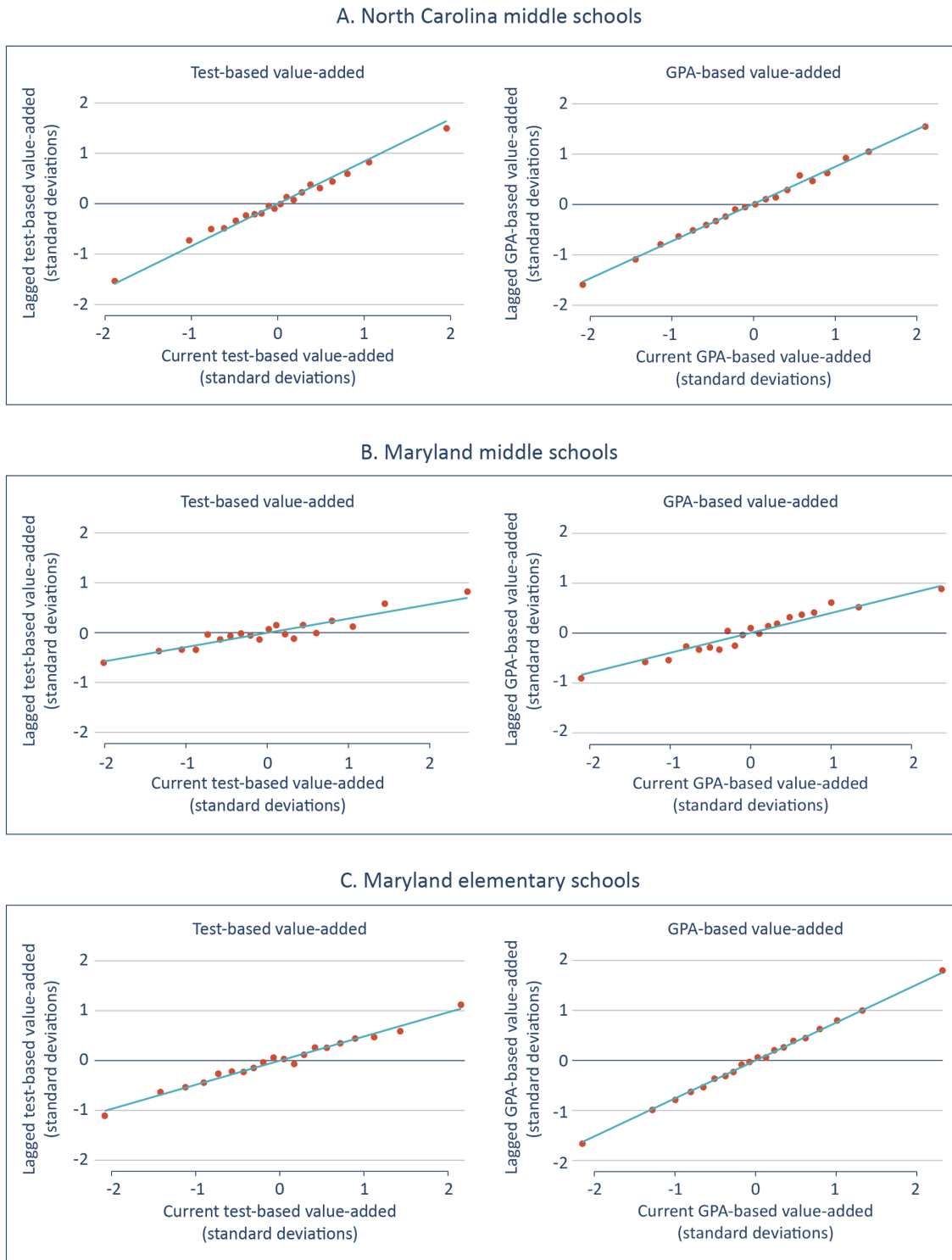
Notes: This figure shows the correlation between GPA- and test-based value-added for North Carolina and Maryland elementary and middle schools.

Importantly, the weakness of these relationships is not attributable to the instability of the measures themselves. In fact, both GPA- and test-based value-added are quite stable over time (Figure 4). For example, in North Carolina middle schools, the intertemporal correlation coefficient (a measure of year-to-year stability) is 0.74 for GPA value-added and 0.86 for test-based value-added (Panel A). And in Maryland, the equivalent figures are 0.40 and 0.29 for middle schools (Panel B) and 0.68 and 0.27 for elementary schools (Panel C). In other words, the Maryland results suggest that GPA value-added is actually more stable than test-based value-added.

Collectively, these strong and statistically significant correlations suggest that the models are capturing something “real” and reasonably consistent about schools’ ability to improve students’ subsequent GPAs and tests scores. Moreover, per the figures (which are approximately linear), this stability is similar across the GPA and test-based value-added distributions. In other words, the results suggest that schools that exhibit a particular form of effectiveness tend to be similarly effective year in and year out, while those that are ineffective tend to remain so.

Ultimately, the fact that these two relatively stable measures of school quality are so weakly correlated with one another is consistent with the notion that test-score and GPA value added capture distinct dimensions of school quality.

Figure 4: Like test-based value-added measures, GPA value-added measures are fairly stable.



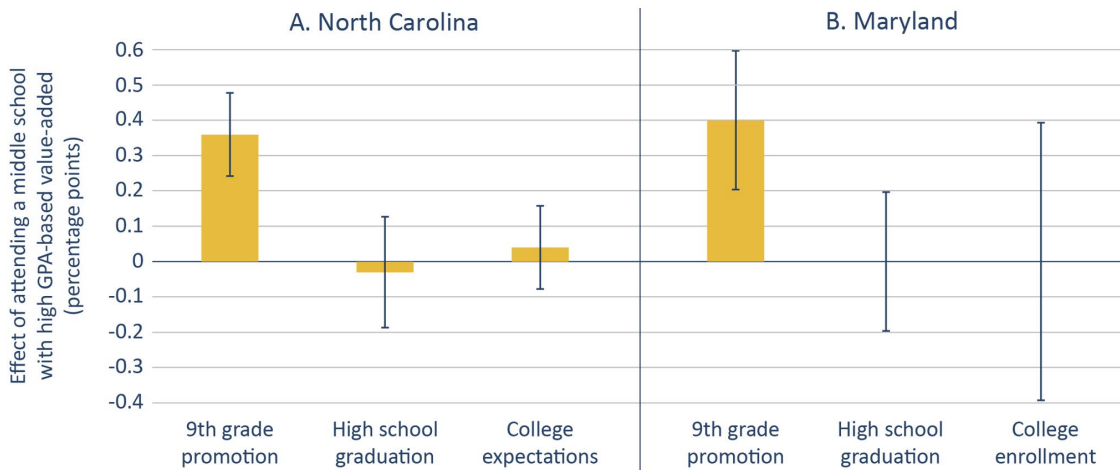
Notes: This figure shows the intertemporal correlation of North Carolina and Maryland elementary and middle schools' value-added to 6th and 9th grade math achievement and GPA. A stronger correlation indicates greater year-to-year stability, which is one indication of reliability.

### Finding 3: Even schools that boost short-run outcomes don't necessarily boost long-run outcomes like high school graduation and college going.

Consistent with Finding 1, enrolling in a middle school with higher GPA value-added significantly reduces the likelihood that a student repeats 9th grade. Specifically, a one standard deviation increase in a middle school's GPA value added in either North Carolina or Maryland reduces the likelihood that a student repeats ninth grade by about 0.4 percentage points – or approximately 8 percent (Figure 5). Like the estimates that are the basis for Finding 1, these effects are robust to controlling for the middle school's test-score value-added, suggesting (once again) that the GPA and test-score value-added measures are identifying different dimensions of school quality.

Yet, despite these results, it's not clear that attending a middle school with high GPA-value-added has any effects on longer-term attainment measures (Figure 5). Indeed, a one SD increase in middle school GPA value-added has no discernible effect on high school graduation or students' college-going expectations in North Carolina (Panel A). Similarly, the long-term impacts of attending a Maryland middle school with high GPA value-added on high school graduation and on-time college enrollment are null (Panel B).

**Figure 5: Despite reducing 9th grade retention, attending a middle school with high GPA value-added has no discernible effect on high school graduation or college going.**



Notes: This figure shows the estimated effect of attending a middle school with a GPA value-added score that is one standard deviation above the average middle school's score on 9th grade retention, high school graduation, and college expectations/college enrollment in North Carolina and Maryland.

Though somewhat discouraging from a policy perspective, these results don't necessarily invalidate the measures given the increasingly high rates of high school graduation, the somewhat imprecise nature of the estimates, and the number of confounding variables that could potentially intervene before graduation and college going decisions.

## Implications

Collectively, the results suggest that measures of “GPA value added” may be fruitfully used as a supplement to measures of “test score value added” to gauge elementary and middle school quality. Although we cannot identify specific mechanisms, it seems likely that GPA value-added measures identify schools that help develop hard-working, engaged, and well-rounded students. Like the skills that are captured by test scores, these traits are valued in the next schools that students attend, as well as in higher education and the workplace. And given the almost non-existent correlation between the two measures, using GPA value-added to capture and report the development of these essential skills has the potential to provide policymakers, parents, and other stakeholders with valuable additional information about school quality that is missing when only test-score value added is available, in addition to encouraging schools to focus on a broader range of content and skills.

Ultimately, using GPA-based growth in the way defined in this report assures school staff that their mission isn’t just proficiency on testing day or uncontroversial promotion to the next grade level. Rather, the mission of any good school is to prepare students for *future* success. GPA value-added accomplishes this in two ways: First, by holding schools accountable for students’ success at their *next* school, it explicitly disincentivizes teaching to the test or socially promoting students in ways that yield short-lived and ultimately illusory gains. Second, it provides a holistic measure of students’ performance that reflects their achievement, engagement, and effort across *all* subjects—not just reading and math.

That’s a test of performance, commitment, and character that is worth teaching to.

## Appendix A: Methodology

We follow current best practice for the estimation of school effects, which proceeds in two steps.<sup>[32]</sup> First, we estimate middle-school GPA value-added scores ( $\theta$ ) by modeling a student's 9th grade GPA (GPA9) as a function of their individual characteristics ( $X$ ), which includes their lagged 8th grade math and reading scores, 8th grade GPA (when available), and indicators of their socio-demographic background; middle school characteristics ( $Z$ ), which are essentially the school-by-year means of the elements of  $X$ ; a middle-school fixed effect ( $\theta$ ); a high-school fixed effect ( $\omega$ ); and an idiosyncratic error ( $\epsilon$ ). Formally, we estimate models of the form

$$GPA9_{ijt} = \alpha X_{ij,t-1} + \beta Z_{m,t-1} + \theta_m + \omega_j + \epsilon_{ijt} \quad (1)$$

where  $i$  indexes students,  $j$  indexes high schools,  $m$  indexes middle schools, and  $t$  indexes years. We follow the same logic to estimate test-score value-added, replacing GPA9 with 9th-grade end-of-course standardized math scores in equation (1). To estimate elementary-school value added we replace the grade 9 and grade 8 measures in equation (1) with analogous grade 6 and 5 measures, respectively. The school value-added measures, i.e., the estimates of  $\theta$ , derived from equation (1) are central to all subsequent analyses. Technically, we compute year-specific, leave-one-out estimates of  $\theta$  that are based on all years except for the current year, to avoid the mechanical endogeneity of a student's outcome being part of their school's value added.<sup>[33]</sup>

Second, we take these GPA and test-score leave-one-out value-added estimates and include them as inputs in models that predict current and future academic performance. These models look similar to those described by equation (1), except that the school FE is replaced by the estimated leave-one-out ( $-t$ ) value-added score(s). Formally, we estimate models of the form:

$$y_{ijt} = \alpha X_{ij,t-1} + \beta Z_{ijt} + \hat{\theta}_{m,-t} + \omega_j + \epsilon_{ijt} \quad (2)$$

where  $y$  is an outcome such as grade-9 end-of-course math scores, grade-9 GPA, or indicators for repeating grade 9, graduating high school, planning to attend college, or college enrollment. The inclusion of estimated regressors in equation (2) make the usual cluster-robust standard errors invalid, so we instead use block-bootstrapped standard errors that account for the imprecision inherent in the estimated school effects; in practice, this decision does not matter, nor does the level of clustering. We probe the robustness of the estimates from equation (2) in several ways: we consider alternative calculations of GPA, including weighted and unweighted, averaging across two years rather than one, and using only core courses. The results are robust to these definitional decisions.



## Appendix B: Additional Tables

**Table B1. Summary Statistics in Maryland**

	<u>Middle School Sample</u>				<u>Elementary School Sample</u>			
	<u>Student level</u>		<u>School-year level</u>		<u>Student level</u>		<u>School-year level</u>	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
GPA_9th Grade (Standardized)	-0.000	(1.00)	-0.304	(0.69)	0.000	(1.00)	-0.112	(0.64)
Weighted GPA_9th Grade (Standardized)	-0.000	(1.00)	-0.303	(0.69)	0.000	(1.00)	-0.110	(0.63)
Math Score_9th Grade	-0.000	(1.00)	-0.126	(0.59)	-0.000	(1.00)	-0.106	(0.55)
On Time High School Graduation	0.846		0.794	(0.17)	0.485		0.482	(0.37)
On Time College Enrollment	0.587		0.511	(0.22)	0.203		0.194	(0.19)
Female	0.495		0.487	(0.12)	0.491		0.490	(0.09)
Economically Disadvantaged	0.398		0.499	(0.28)	0.446		0.509	(0.29)
English Language Learner	0.069		0.059	(0.11)	0.127		0.118	(0.17)
Special Education	0.146		0.165	(0.10)	0.149		0.160	(0.08)
Homelessness	0.011		0.014	(0.03)	0.012		0.015	(0.03)
Asian	0.071		0.048	(0.07)	0.069		0.054	(0.08)
Black	0.319		0.459	(0.35)	0.319		0.390	(0.34)
Hispanic	0.149		0.123	(0.16)	0.168		0.157	(0.19)
Missing Race	0.000		0.000	(0.00)	0.000		0.000	(0.00)
Indian/Alaska	0.002		0.003	(0.01)	0.003		0.003	(0.01)
Observations	286958		1873		288368		4686	

**Table B2. Summary Statistics in North Carolina**

	<u>Student level</u>		<u>School-year level</u>	
	Mean	SD	Mean	SD
Standardized GPA	0.01	(0.98)		
Standardized Math-1 Score	0.02	(0.96)		
Repeated Grade 9	0.04			
HS Diploma	0.86			
Plans to Attend 4-Year College	0.05			
School GPA VAM	-0.00	(1.00)	-0.01	(0.98)
School Math Score VAM	0.00	(1.00)	-0.00	(0.93)
Grade 8 Math z score	0.04	(0.99)	-0.04	(0.45)
Grade 8 Reading z score	-0.01	(0.99)	-0.07	(0.41)
Male	0.51		0.51	(0.07)
Economically Disadvantaged	0.46		0.50	(0.22)
English Language Learner	0.04		0.04	(0.06)
Gifted	0.18		0.15	(0.12)
Special Education	0.11		0.13	(0.08)
Asian	0.03		0.02	(0.04)
Black	0.25		0.26	(0.24)
Hispanic	0.15		0.14	(0.12)
American Indian	0.01		0.02	(0.09)
Multi-race	0.04		0.04	(0.03)
White	0.53		0.53	(0.28)

**Table B3. Variation in Test Score and GPA School-VAM Short-Run Effects**

<b>North Carolina Middle Schools</b>						
Outcome	Grade 9 Standardized GPA (unweighted)			Grade 9 Standardized Math 1 Score		
GPA VAM (ms)	0.0929*** (0.0028)		0.0936*** (0.0027)	-0.0005 (0.0038)		-0.0083*** (0.0030)
Math score VAM (ms)		-0.0011 (0.0043)	-0.0100** (0.0041)		0.1275*** (0.0134)	0.1283*** (0.0135)
Observations	618,586	618,134	618,134	400,987	400,956	400,956
R-squared	0.5496	0.5425	0.5495	0.5438	0.5516	0.5517
<b>Maryland Middle Schools</b>						
Outcome	Grade 9 Standardized GPA			Grade 9 Standardized Math Score		
GPA VAM	0.054*** (0.004)		0.058*** (0.005)	0.006 (0.007)		0.001 (0.005)
Math Score VAM		0.001 (0.003)	-0.001 (0.004)		0.117*** (0.010)	0.116*** (0.011)
Observations	273961	273959	273959	155999	155997	155997
R-squared	0.525	0.522	0.522	0.422	0.431	0.430
<b>Maryland Elementary Schools</b>						
Outcome	Grade 6 Standardized GPA			Grade 6 Standardized Math Score		
GPA VAM	0.089*** (0.010)		0.089*** (0.011)	-0.004 (0.004)		-0.006 (0.004)
Math Score VAM		0.023*** (0.006)	0.022*** (0.006)		0.025*** (0.003)	0.027*** (0.003)
Observations	219915	219915	220295	215538	215538	215916
R-squared	0.468	0.461	0.446	0.752	0.753	0.747

**Table B4. Variation in Test Score and GPA School-VAM Long-Run Effects**

<b>Repeat 9<sup>th</sup> grade</b>						
	North Carolina			Maryland		
GPA VAM	-0.0035*** (0.0006)		-0.0036*** (0.0006)	-0.004** (0.0010)		-0.004** (0.0010)
Math Score VAM		0.0001 (0.0006)	0.0005 (0.0006)		-0.003* (0.0010)	-0.003** (0.0010)
Observations	618,586	618,134	618,134	275,189	275,187	275,187
R-squared	0.091	0.0908	0.0911	0.171	0.171	0.169

<b>High School Graduation</b>						
	North Carolina			Maryland		
GPA VAM	-0.0002 (0.0008)		-0.0003 (0.0008)	-0.000 (0.001)		-0.000 (0.001)
Math Score VAM		0.0019* (0.0010)	0.0020** (0.0010)		-0.002 (0.001)	-0.002 (0.001)
Observations	618,586	618,134	618,134	275,189	275,187	275,187
R-squared	0.6932	0.6931	0.6931	0.129	0.129	0.129

<b>Expect/Enroll in College</b>						
	North Carolina			Maryland		
GPA VAM	0.0005 (0.0006)		0.0004 (0.0006)	-0.000 (0.002)		-0.000 (0.002)
Math Score VAM		0.0014** (0.0007)	0.0014** (0.0007)		-0.003** (0.001)	-0.003** (0.001)
Observations	618,586	618,134	618,134	275,189	275,187	275,187
R-squared	0.0868	0.0869	0.0869	0.282	0.282	0.282

## Endnotes

<sup>[1]</sup> Jackson, C. Kirabo. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes." *Journal of Political Economy* 126, no. 5 (2018): 2072–107. Jackson, C. Kirabo, et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508.

<sup>[2]</sup> Bowen, William G., Matthew M. Chingos, and Michael McPherson. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press, 2009. Roderick, Melissa, et al. *From High School to the Future: A First Look at Chicago Public School Graduates' College Enrollment, College Preparation, and Graduation from Four-Year Colleges*. Chicago, IL: University of Chicago Consortium on Chicago School Research, 2006. <https://consortium.uchicago.edu/sites/default/files/2018-10/Postsecondary.pdf>. Bowers, A. J., R. Sprott, and S. A. Taff. "Do We Know Who Will Drop Out? A Review of the Predictors of Dropping Out of High School: Precision, Sensitivity, and Specificity." *The High School Journal* 96, no. 2 (2013): 77–100. Brookhart, Susan M., et al. "A Century of Grading Research: Meaning and Value in the Most Common Educational Measure." *Review of Educational Research* 86, no. 4 (2016): 803–48.

<sup>[3]</sup> Ibid.

<sup>[4]</sup> Note that a school's GPA is the average student's GPA across all subjects including untested subjects. Results are similar when we consider different versions of GPA, such as using only core subjects or weighted GPA by number of credits.

<sup>[5]</sup> Because ESSA mandates that accountability systems include data for all students as well as specific subgroups, the measures that are the basis for this study may not be ESSA compliant; however, it is our belief that a suitably motivated state could construct analogous indicators for students who take other math courses in 9th grade, with the goal of constructing a more ESSA-compliant index.

<sup>[6]</sup> Easton, John Q., Esperanza Johnson, and Lauren Sartain. *The Predictive Power of Ninth-Grade GPA*. Chicago, IL: University of Chicago Consortium on School Research, 2017. <https://uchicago-dev.geekpak.com/sites/default/files/2018-10/Predictive%20Power%20of%20Ninth-Grade-Sept%202017-Consortium.pdf>.

<sup>[7]</sup> Bowen, William G., Matthew M. Chingos, and Michael McPherson. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press, 2009. Roderick, Melissa, et al. *From High School to the Future: A First Look at Chicago Public School Graduates' College Enrollment, College Preparation, and Graduation from Four-Year Colleges*. Chicago, IL: University of Chicago Consortium on Chicago School Research, 2006. <https://consortium.uchicago.edu/sites/default/files/2018-10/Postsecondary.pdf>. Bowers, A. J., R. Sprott, and S. A. Taff. "Do We Know Who Will Drop Out? A Review of the Predictors of Dropping Out of High School: Precision, Sensitivity, and Specificity." *The High School Journal* 96, no. 2 (2013): 77–100. Brookhart, Susan M., et al. "A Century of Grading Research: Meaning and Value in the Most Common Educational Measure." *Review of Educational Research* 86, no. 4 (2016): 803–48.

<sup>[8]</sup> Note that ninth grade is often considered a "make or break" year for students, which is why some states and districts use the number of classes students fail in ninth grade as an early warning system.

<sup>[9]</sup> For more, see <https://fordhaminstitute.org/national/research/imperfect-attendance-toward-fairer-measure-student-absenteeism>

<sup>[10]</sup> Consistent with the informal standard that has been established for test-based teacher value-added measures, we propose that GPA-based value-added measures be reported when the grades of at least five graduates of an elementary or middle school can be compared.

<sup>[11]</sup> There are at least two reasons for this: First, the law requires that all public schools be evaluated in the same way, but the proposed measure may not work for high schools with a single feeder middle school. Second, the fact that the measure that is the basis for this report excludes about two fifths of students is potentially inconsistent with ESSA's requirement that all students be included (though as noted, it is probably feasible to construct a version of the measure that includes a higher percentage of students). In practice, there is probably some flexibility on both counts.

<sup>[12]</sup> Jennings, Jennifer L., and Jonathan Marc Bearak. "Teaching to the Test' in the NCLB Era: How Test Predictability Affects Our Understanding of Student Performance." *Educational Researcher* 43, no. 8 (2014): 381–89. Koedel, Cory, Rebecca Leatherman, and Eric Parsons. "Test Measurement Error and Inference from Value-Added Models." *The BE Journal of Economic Analysis & Policy* 12, no. 1 (2012). Park, R. Jisung, Joshua Goodman, Michael Hurwitz, and Jonathan Smith. "Heat and Learning." *American Economic Journal: Economic Policy* 12, no. 2 (2020): 306–39.

<sup>[13]</sup> Liu, Jing. *Imperfect Attendance: Toward a Fairer Measure of Student Absenteeism*. Washington, DC: Thomas B. Fordham Institute, 2022.

<sup>[14]</sup> Gershenson, Seth. "Linking Teacher Quality, Student Attendance, and Student Achievement." *Education Finance and Policy* 11, no. 2 (2016): 125–49. Liu, Jing, and Susanna Loeb. "Engaging Teachers: Measuring the Impact of Teachers on Student Attendance in Secondary School." *Journal of Human Resources* 56, no. 2 (2021): 343–79.

<sup>[15]</sup> Bowen, William G., Matthew M. Chingos, and Michael McPherson. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press, 2009. Bowers, A. J., R. Sprott, and S. A. Taff. "Do We Know Who Will Drop Out? A Review of the Predictors of Dropping Out of High School: Precision, Sensitivity, and Specificity." *The High School Journal* 96, no. 2 (2013): 77–100. Roderick, Melissa, et al. *From High School to the Future: A First Look at Chicago Public School Graduates' College Enrollment, College Preparation, and Graduation from Four-Year Colleges*. Chicago, IL: University of Chicago Consortium on Chicago School Research, 2006. <https://consortium.uchicago.edu/sites/default/files/2018-10/Postsecondary.pdf>. Friedman, J., B. Sacerdote, and M. Tine. *Standardized Test Scores and Academic Performance at Ivy-Plus Colleges*. Opportunity Insights, 2024. [https://opportunityinsights.org/wp-content/uploads/2024/01/SAT\\_ACT\\_on\\_Grades.pdf](https://opportunityinsights.org/wp-content/uploads/2024/01/SAT_ACT_on_Grades.pdf).

<sup>[16]</sup> Deming, David J. "Better Schools, Less Crime?" *The Quarterly Journal of Economics* 126, no. 4 (2011): 2063–115. Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. "School Choice, School Quality, and Postsecondary Attainment." *American Economic Review* 104, no. 3 (2014): 991–1013. Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508. Beuermann, Diether W., C. Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. "What Is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output." *The Review of Economic Studies* 90, no. 1 (2023): 65–101. Beuermann, Diether W., and C. Kirabo Jackson. "The Short-and Long-Run Effects of Attending the Schools That Parents Prefer." *Journal of Human Resources* 57, no. 3 (2022): 725–46.

<sup>[17]</sup> Jackson et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508. Liu,

Jing. *Imperfect Attendance: Toward a Fairer Measure of Student Absenteeism*. Washington, DC: Thomas B. Fordham Institute, 2022.

[18] Rockoff, J.E. (2004); Harris, D. N. (2011).

[19] Gershenson, Seth. "Linking Teacher Quality, Student Attendance, and Student Achievement." *Education Finance and Policy* 11, no. 2 (2016): 125–49. Liu, Jing, and Susanna Loeb. "Engaging Teachers: Measuring the Impact of Teachers on Student Attendance in Secondary School." *Journal of Human Resources* 56, no. 2 (2021): 343–79. Jackson, C. Kirabo. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes." *Journal of Political Economy* 126, no. 5 (2018): 2072–107. Jackson et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508.

[20] Loeb, Susanna, Matthew S. Christian, Heather Hough, Robert H. Meyer, Andrew B. Rice, and Martin R. West. "School Differences in Social-Emotional Learning Gains: Findings from the First Large-Scale Panel Survey of Students." *Journal of Educational and Behavioral Statistics* 44, no. 5 (2019): 507–42. Jackson et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508. Liu, Jing. *Imperfect Attendance: Toward a Fairer Measure of Student Absenteeism*. Washington, DC: Thomas B. Fordham Institute, 2022.

[21] Jackson, C. Kirabo. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes." *Journal of Political Economy* 126, no. 5 (2018): 2072–107.

[22] Paunesku, David, Gregory M. Walton, Carissa Romero, Eric N. Smith, David S. Yeager, and Carol S. Dweck. "Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement." *Psychological Science* 26, no. 6 (2015): 784–93. <https://doi.org/10.1177/0956797615571017>.

[23] For an illustrative synthesis, see Deming, David J., and Mikko I. Silliman. "Skills and Human Capital in the Labor Market." *NBER Working Paper* No. 32908, September 2024. <https://www.nber.org/papers/w32908>.

[24] Guskey, Thomas R., and Laura J. Link. "Exploring the Factors Teachers Consider in Determining Students' Grades." *Assessment in Education: Principles, Policy & Practice* 26, no. 3 (2019): 303–20.

[25] Easton, John Q., Esperanza Johnson, and Lauren Sartain. *The Predictive Power of Ninth-Grade GPA*. Chicago, IL: University of Chicago Consortium on School Research, 2017. <https://uchicago-dev.geekpak.com/sites/default/files/2018-10/Predictive%20Power%20of%20Ninth-Grade-Sept%202017-Consortium.pdf>.

[26] Allensworth, Elaine M., and Kallie Clark. "High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions about Consistency across High Schools." *Educational Researcher* 49, no. 3 (2020): 198–211. <https://doi.org/10.3102/0013189X20902110>.

[27] Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (2014): 2593–632. Jackson et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508.

[28] Note that this test-based value-added measure, which is based on students' subsequent math achievement, differs from the value-added measures that are typically used in state accountability systems, which are based on students' math achievement at the end of the relevant school year.

<sup>[29]</sup> It is also robust to whether GPA is measured over one or two years (i.e., grades 9 and 10), and overall or just core classes.

<sup>[30]</sup> Jackson et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508.

<sup>[31]</sup> C. Kirabo Jackson, "What do test scores miss? The importance of teacher effects on non-test score outcomes," *Journal of Political Economy* 126, no. 5 (2018): 2072–107; Backes, Ben, et al. "How to measure a teacher: The influence of test and non-test value-added on long-run student outcomes." *Journal of Human Resources* (2024).

<sup>[32]</sup> Jackson et al. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights* 2, no. 4 (2020): 491–508.

<sup>[33]</sup> Chetty et al. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (2014): 2593–632.