

USING PUBLIC ACCOUNTABILITY DATA TO PROMOTE EQUITY
IN MICHIGAN SCHOOL DISTRICTS

By

Jennifer A. Gruber

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology—Doctor of Philosophy

2023

ABSTRACT

School accountability is the primary method the United States public education system uses to monitor the quality of local and state education systems and promote positive educational outcomes. The current accountability system under the *Every Student Succeeds Act of 2015* (ESSA) grants states the autonomy to design their own policies and metrics for school and district performance. Researchers and educators have raised concerns about these accountability systems, including their approach to identifying schools and districts that need improvement, their potential harmful consequences, and their lack of attention to the structural causes of educational inequities. School is only one system—albeit an impactful, important one—within a student’s social ecology. Schools with lower performance are often situated within contexts that perpetuate inequities and limit their ability to respond to the barriers their students face.

Using two sources of publicly available education data that report various student-, school-, and district-level characteristics (MI School Data and Civil Rights Data Collection), I conducted an exploratory study of schools in 12 public school districts that—as of September 2021—had a partnership agreement with the Michigan Department of Education (i.e., were the focus of state-level intervention under the current Michigan school accountability system). Specifically, I used multilevel modeling to examine school- and district-level measured indicators of structural factors (e.g., school staff-to-student ratios; district finances) and student achievement (e.g., test scores) and disciplinary (e.g., suspensions) outcomes and their relations over time in schools in these 12 school districts relative to a matched comparison sample. I also incorporated an explicit focus on equity by examining the extent to which these relationships differed across student subgroups by race and ethnicity, socioeconomic status, and disability status.

My primary aims were to examine the extent to which: (1) partnership district schools differed from matched comparison district schools on student outcomes over time; (2) proxies for structural factors (e.g., enrollment, financial status) impacted student outcomes; and (3) partnership district schools differed from matched comparison district schools in terms of equity.

For my first aim, I found that partnership district schools had worse average academic outcomes than matched comparison district schools, but the differences between the schools were stable throughout the years of data included in my study. Given the stability of these differences, comprehensive school reform or community-level supports might be the best approach to address deeply rooted barriers faced by schools. For my second aim, I found that several structural factors (e.g., student mobility, the enrollment of historically marginalized students) accounted for academic outcomes over and above school accountability metrics. Given the potential consequences schools face if they do not meet the specific goals outlined in the agreements, it is important to consider how data on these structural factors could be leveraged to identify areas to best support schools or to account for factors outside of a school's control. Less clear patterns emerged for disciplinary outcomes, which might be an important area for future research and consideration. For my third aim, I was only able to examine differences across student subgroups for one outcome (math growth percentiles). I found that all student subgroups except Latine students had worse math growth percentiles in partnership district schools compared to matched comparison district schools, but few structural factors emerged as statistically significant to explain these differences. Overall, my findings suggest specific areas of promise for Michigan and other states to better align their school accountability systems with ESSA's goals of providing an equitable, holistic education.

This dissertation is dedicated to the many who kept me laughing along the way.

To my parents, Diana and Erich.

To my siblings, Linnea, Ryan, and Cassie.

To my forever friends, Alysia and Rachael.

To my graduate school family, in particular, MK, Isi, Funmi, and Sara.

To my advisor, Ignacio.

Finally, to my home, Ryan, Hagrid, and Snoops.

“I had the epiphany that laughter was light, and light was laughter, and that this was the secret of the universe.”

— Donna Tartt, *The Goldfinch*

TABLE OF CONTENTS

INTRODUCTION	1
LITERATURE REVIEW	3
School Accountability in the US	3
<i>Every Student Succeeds Act</i>	3
<i>Considerations for School Accountability</i>	6
Ecological Perspectives on School Accountability	11
<i>Demographic Characteristics and Educational Outcomes</i>	13
<i>School and District Resources</i>	16
THE MICHIGAN CONTEXT	18
School Accountability in Michigan	18
<i>School Quality Index Scores</i>	18
<i>Tiered Support Systems</i>	20
Considerations for Michigan School Accountability	21
<i>SQI Scores</i>	21
<i>Potential Consequences of Michigan Accountability Systems</i>	23
<i>Ecological Considerations for Students and Districts in Michigan</i>	24
THE CURRENT PROJECT	26
Research Questions and Hypotheses	27
METHODS	29
Data Sources	29
<i>MI School Data</i>	29
<i>Civil Rights Data Collection</i>	30
Sampling	30
<i>Matching Procedure</i>	30
Sample	33
Variables and Constructs	36
<i>Measured Indicators of Structural Factors</i>	36
<i>Student Outcomes</i>	39
<i>School and District Characteristics</i>	42
Missing Data	43
Outliers	45
Data Preparation	47
Data Reduction using Confirmatory Factor Analysis	48
<i>CFA Model Fit Indices</i>	48
<i>CFA Results</i>	49
Multilevel Models	52
RESULTS	57
Descriptive Statistics	57
<i>Measured Indicators of Structural Factors</i>	57

<i>Student Outcomes</i>	60
<i>School Characteristics</i>	63
RQ1: How do Partnership Districts—relative to their Matched Comparisons—experience changes in student outcomes over time?	64
<i>Attendance</i>	64
<i>Assessment</i>	66
<i>Discipline</i>	75
<i>Summary of RQ1 Findings</i>	82
RQ2: How do proxies for structural factors predict student outcomes?	82
<i>Attendance</i>	82
<i>Assessment</i>	85
<i>Discipline</i>	96
<i>Summary of RQ2 Findings</i>	105
RQ3: Are the patterns identified in RQ1 and RQ2 the same when examining outcomes for different groups of students?	107
<i>Black Students</i>	107
<i>Latine Students</i>	109
<i>White Students</i>	111
<i>Students who Qualify for FRL</i>	113
<i>Students with an IEP</i>	116
<i>Summary of RQ3 Findings</i>	118
DISCUSSION	121
RQ1: How do partnership district schools—relative to their matched comparisons—experience changes in student outcomes over time?	121
<i>Stability of Student Outcomes Over Time</i>	122
<i>Differences in Demographic Characteristics of Partnership Districts and Matched Comparison Districts</i>	124
<i>Student Outcomes Not Predicted by Partnership Status</i>	125
RQ2: How do proxies for structural factors predict student outcomes?	127
<i>Overall Findings</i>	127
<i>Student Enrollment Characteristics</i>	129
<i>Student Mobility Characteristics</i>	131
<i>School and District Resources</i>	133
RQ3: Are the patterns identified in RQ1 and RQ2 the same when outcomes for different groups of students?	135
<i>Limitations</i>	138
CONCLUSION	142
REFERENCES	144
APPENDIX A: LIST OF VARIABLES AND SOURCES	155
APPENDIX B: PERFORMANCE LEVELS FOR THE M-STEP	158
APPENDIX C: SUPPLEMENTAL TABLES	159

INTRODUCTION

In the United States (US), there are a multitude of positive outcomes associated with educational attainment. More education is associated with less risk of cardiovascular disease, less disability, and a greater life expectancy (Kubota et al., 2017; Laditka & Laditka, 2016).

Education is also positively associated with life satisfaction, happiness, and income (Assari, 2019; Greenstone et al., 2013; Lawless & Lucas, 2011). Further, counties with higher levels of educational attainment experience lower poverty rates, even when accounting for other county-level demographics such as age, race, gender, and family structure (Levernier et al., 2000).

Given the significant health and economic benefits of educational attainment, the US places a high value on education. Through financial and political investments, educational attainment in the US has grown substantially over the last century (Harris & Herrington, 2006; Pfeffer & Hertel, 2015). In 1940, approximately 39% of individuals 25 years or older graduated high school in the US; by 2020, this proportion increased to approximately 96% (US Census Bureau, 2021). As of 2017, there were nearly 60 million students (pre-kindergarten through 12th grade) enrolled in over 130,000 schools in the US (Hussar et al., 2020). Moreover, US students' math and reading proficiency rates have increased over the last 30 years (Hussar et al., 2020).

However, on a global scale the US education system remains in the middle of the pack in terms of improving student achievement. Compared to 48 other nations, the US has shown worse growth in reading, math, and science assessments than 24 countries over a 14-year period (from 1995 to 2009). Eleven countries have growth rates *twice* the size of the US (Hanushek et al., 2012). Additionally, in the US educational inequities persist based on social characteristics or conditions such as locale, class, race and ethnicity, and gender (Bauer et al., 2018; Hochschild & Shen, 2014; Hussar et al., 2020; Pfeffer & Hertel, 2015; Xie et al., 2015). For example, Pfeffer

and Hertel (2015) found educational attainment in the US has substantially increased over the last 80 years but continues to be strongly related to social class (e.g., a parent's educational attainment predicts their children's educational attainment).

School accountability is one of the key approaches used in the US to promote quality education (Bae, 2018). Briefly, school accountability systems encompass policies that require schools to meet certain performance benchmarks; if schools are not meeting these benchmarks, local or state education systems are required to intervene (Adler-Greene, 2019; Bae, 2018). In this exploratory study, I examined student outcome trends in schools and districts that Michigan's accountability system flagged as requiring the most intense state-level intervention. In this literature review, I provide a review of school accountability systems in the US and outline several important considerations for these systems, including how school quality is measured and the potential consequences for schools under these systems. I then apply an ecological framework to school accountability to consider the underlying structural barriers that might lead to students and schools not meeting their performance benchmarks. Finally, I put this information into the Michigan context.

LITERATURE REVIEW

School Accountability in the US

The US education system implements accountability structures at the local, state, and federal levels to ensure schools provide high quality educational experiences to students. These accountability systems operate under the assumption that incentives—whether positive (e.g., financial awards) or negative (e.g., funding restrictions)—and public pressure (e.g., publicizing school data and rankings) will motivate school, district, and state leaders to improve student outcomes (Bae, 2018). Current approaches to school accountability are governed by federal legislation, the *Every Student Succeeds Act of 2015*.

Every Student Succeeds Act

The *Every Student Succeeds Act of 2015* (ESSA) provides guidance and requirements for: (1) measuring school quality; (2) tracking student progress; (3) identifying schools that need assistance; (4) providing a well-rounded education; and (5) developing school improvement plans (Adler-Greene, 2019; Bae, 2018; Darling-Hammond et al., 2016). Under ESSA, states establish and track their own measures of school quality or progress using multiple, weighted indicators following certain requirements (McGuinn, 2016). For consistency, throughout this document I refer to these composite measures as “school quality measures”, which are required to include: state assessment scores for math and language arts, student growth in math and English Language Arts (ELA), graduation rate, English Language Learner (ELL) students’ English proficiency, and at least one additional measure of school quality or student success (Cardichon & Darling-Hammond, 2017; Darling-Hammond et al., 2016; McGuinn, 2016).

ESSA also requires that the academic achievement indicators be of more weight than other indicators (Cardichon & Darling-Hammond, 2017; Darling-Hammond et al., 2016;

McGuinn, 2016). States are required to track and report these outcomes for the general student population and for student subgroups, such as groups broken down by student race and ethnicity, disability status, or socioeconomic status (Darling-Hammond et al., 2016; McGuinn, 2016). ESSA also incorporates an explicit focus on ELL students, requiring states to track and hold schools accountable to their English proficiency progress, as well as requiring states to outline how they will provide resources for these students in their support plans (Adler-Greene, 2019).

ESSA prescribes an incremental approach to targeting schools that are not meeting their goals with three categories, from most to least intensive: (1) Comprehensive Support; (2) Targeted Support; and (3) Additional Targeted Support (McGuinn, 2016). Comprehensive Support schools are public schools that receive Title I funds and are either in the bottom five percent of schools—per the metrics established by the state—or fail to graduate one third or more of their students (Darling-Hammond et al., 2016; McGuinn, 2016). The districts to which such schools belong are required to create and implement comprehensive support plans to address the areas that need improvement (Darling-Hammond et al., 2016). Targeted and Additional Targeted Support schools are public schools that receive Title I funds and have one or more student subgroups performing in the bottom five percent, relative to the rest of the state (McGuinn, 2016). ESSA requires states to identify schools that need additional supports and create support plans every three years, granting the states autonomy over determining the supports and/or consequences for these schools (Darling-Hammond et al., 2016).

ESSA incentivizes schools to provide a well-rounded education by requiring states to develop and implement plans that address issues of school climate and requiring districts to use a set proportion of federal funds towards resources (e.g., school counselors) or extracurricular activities (Adler-Greene, 2019; Bae, 2018). Further, ESSA recommends that the support plans

developed by the state outline how schools will reduce exclusionary discipline practices (i.e., out-of-school suspension, expulsion) and provides awards for districts or states that aim to transform school climate and discipline by shifting to focus on restorative practices (Adler-Greene, 2019; McGuinn, 2016).

ESSA replaced the *No Child Left Behind Act of 2001* (NCLB). NCLB aimed to improve student achievement and address documented achievement inequities by: (1) ensuring all teachers and instructional staff are “highly qualified”; (2) promoting the use of evidence-based practices in schools; and (3) increasing caregiver involvement in their students’ education (Simpson et al., 2004). NCLB set ambitious goals, aiming to have all students in the US meet proficiency standards by 2014 (Kim & Sunderman, 2005). In NCLB the outcomes that defined student achievement—and thus became the primary indicators of school and district performance—were standardized test scores on reading and math achievement, with a later addition of standardized test scores on science achievement (Pederson, 2007).

NCLB established a federally centralized system of school accountability. All states identified benchmarks for expected yearly growth in reading and math scores, known as Adequate Yearly Progress (Kim & Sunderman, 2005; Simpson et al., 2004). NCLB included positive incentives, such as financial awards or public acknowledgements, and sanctions, such as decreased flexibility in use of federal funds (Simpson et al., 2004). Sanctions increased in severity each year a school did not meet their Adequate Yearly Progress, and could be as extreme as school closure or state takeover (Kim & Sunderman, 2005; Mintrop & Sunderman, 2009).

Although there is some evidence that overall proficiency on standardized tests increased during NCLB, the gaps between historically marginalized students’ test scores and their peers’

test scores did not appear to close (Harris & Herrington, 2006; Mintrop & Sunderman, 2009). In fact, in some instances they widened; the gap between standardized test scores for Black students attending racially segregated and low income schools and the national average was even greater after NCLB (Adler-Greene, 2019). The narrow focus on standardized testing, reading, math, and science impacted the prioritization of—and resources allocated to—other subjects and curriculum (Pederson, 2007). Further, the sanctions associated with not meeting Adequate Yearly Progress put immense pressure on educators to increase test scores. Some educators did not want to keep students who struggle with testing in their classroom, and some administrators encouraged the use of self-contained classrooms or transferring special education students to other districts (Adler-Greene, 2019).

ESSA differs from NCLB in key ways. ESSA shifted from NCLB's centralized, uniform accountability system to develop more localized systems that enable states to establish their own benchmarks, incentives, and sanctions (Adler-Greene, 2019; Darling-Hammond et al., 2016). Under ESSA, states determine the specific measures used for school quality or student success and can add additional indicators if they choose (Bae, 2018). ESSA eliminated the Adequate Yearly Progress system and highly qualified educator requirements, shifting its focus to promoting high-quality and well-rounded education rather than just increased test scores (Adler-Greene, 2019; Darling-Hammond et al., 2016).

Considerations for School Accountability

The approach to accountability under ESSA is promising in its flexibility as one of the biggest criticisms of NCLB was its inability to respond to local contexts (Bae, 2018). However, relative to NCLB, ESSA sets fewer federal requirements, granting states increased autonomy in both how they identify schools that need support and how they intervene in these schools. State

accountability systems could easily maintain more traditional systems of accountability, rather than transforming their system to reflect ESSA's new priorities of equity and school transformation (Adler-Greene, 2019; McGuinn, 2016). In fact, as of 2018 only 17 states planned to report outcomes for students with disabilities separately and only 10 detailed strategies to support these students in their plans to implement ESSA (Turner et al., 2018). This means that in 33 states, students with disabilities could not be meeting their performance benchmarks but this would not appear in schools' quality measures or rankings.

Additionally—despite ESSA's emphasis on schools providing a well-rounded education—few states have formally included important non-academic indicators such as socioemotional learning competencies or school climate into their accountability systems (Dusenbury et al., 2018; Schneider et al., 2021). Researchers have found that including these indicators can change a school's accountability rating, and their inclusion might have a particularly positive impact for schools that underperform on traditional accountability metrics and that serve historically marginalized students (Hough et al., 2017; Schneider et al., 2021). There are several additional factors that should be taken into consideration for school accountability. Two that are of direct importance are the validity of approaches to school quality measurement and the potential consequences of flagging schools as needing improvement.

Validity of School Quality Measures. ESSA prescribes a heavy reliance on standardized test scores to measure school quality, and most states continue to use the same standardized tests administered during NCLB (Close et al., 2018; McGuinn, 2016). Researchers have raised concerns regarding the reliability and validity of standardized tests, particularly their strong, consistent relationship with student demographic characteristics (Kane & Staiger, 2002; Schneider et al., 2021). For example, in a nationally representative sample Hegedus (2018) found

that over 50% of a school's median assessment scores were accounted for by the proportion of students that qualified for free-and-reduced lunch in a nationally representative sample. Tienken and colleagues (2017) found that knowing basic community demographics—a community's percentage of individuals making over \$200,000 a year, in poverty, and with bachelor's degrees—predicted middle school math and reading scores. Further, the extent to which improvements on one standardized test are related to improvements on other assessments is unclear (Bae, 2018). For instance, in multiple states, student performance on the state's standardized test was greater than relative performance on the National Assessment of Educational Progress, a low-stakes assessment administered to a representative national sample (Jacob, 2007).

Researchers have raised additional concerns about the validity of school quality measures given that the individual indicators may not be related (Schneider et al., 2021). For example, another required indicator under ESSA is student growth, as defined by each state. Hegedus (2018) found that the bottom five percent of schools in standardized testing had similar score growth trajectories as their higher performing counterparts, indicating these two measures might be independent of each other. In fact, several studies have demonstrated that many schools categorized as “underperforming” (e.g., in the bottom five percent, in the bottom quintile) based on their average standardized test scores would not be considered underperforming based on their ranked growth or growth rate (Downey et al., 2008; Hegedus, 2018). In other words, examining a school's average growth on assessment scores tells a different story than a school's average raw scores, and decisions over the weight that these are given in estimations of school quality can change which schools are flagged as needing intervention. This highlights the instability of these school quality measures and the importance of including valid indicators.

Finally, many indicators have not been systematically examined for inclusion in school quality measures. ESSA allows and encourages states to include additional indicators that reflect holistic aspects of students' educational experiences, such as disciplinary practices (e.g., suspension and expulsion) or school resources (e.g., access to counselors, social workers), but there is little guidance on how to do so (Cardichon & Darling-Hammond, 2017). Additional research is needed to understand the extent to which these non-academic indicators might be related to school quality measures, as well as to determine opportunities for their inclusion in accountability metrics.

Consequences of School Accountability Policies. These concerns about measuring school quality are particularly important given the potential consequences for schools not meeting their state benchmarks or falling into the bottom five percent of schools. Although each state varies in how they approach reforming schools not meeting state benchmarks, some of the more extreme consequences can include school closures (Sunderman et al., 2017). The rationale behind closing a school that is not meeting performance benchmarks is that sending students to higher performing schools will improve their educational experiences and subsequent academic outcomes (Bross et al., 2016; Brummet, 2014). However, when schools are closed there is consistent evidence of a negative impact on communities—given their role as a neighborhood resource hub and association with social and financial capital—and mixed evidence of benefits to students, teachers, or families (Bogart & Cromwell, 2000; Bross et al., 2016; Garnett, 2014; Leyden, 2003; Sunderman et al., 2017; Tieken & Auldrige-Reveles, 2019). For example, Kirshner and colleagues (2010) found that students who transitioned to a new school after a school closure performed worse on reading, math, and writing assessments and had lower graduation rates and higher dropout rates, whereas Brummet (2014) found that students who

transferred from a lower achieving school that closed to a higher achieving school experienced gains in reading and math assessment scores.

Another potential consequence is enrollment decline; many states have school choice policies that allow caregivers to send their students to other schools and districts regardless of their residential status, with varying levels of approval required (Logan, 2018). Some states have school accountability policies that make this process even easier if students are attending a school that is not meeting their performance benchmarks (Bross et al., 2016). For example, multiple states have voucher programs that allow caregivers to apply for a voucher to send their student to a non-residential public school, charter school, or private school, and automatically grant eligibility to students attending schools that are not meeting their performance benchmarks (Carnoy, 2017; Feng et al., 2018). Despite their popularity, there is mixed evidence that these voucher programs improve student achievement, with some studies demonstrating negative outcomes for students that transfer to other schools (Abdulkadiroglu et al., 2015; Carnoy, 2017).

Many public school districts experience enrollment decline greater than overall population decline as caregivers send their students to charter schools or other higher performing public schools (Garnett, 2014). As such, school choice can exacerbate the segregation of schools and districts based on class, race and ethnicity, and special education status as families with more resources have the ability to send their student to another school based on performance (Mordechay & Orfield, 2017; Roda & Wells, 2013; Winters, 2015). For example, Glazerman and Dotter (2017) found that caregivers applying to Washington D.C.'s school choice lottery system were more likely to prefer schools that: served a higher proportion of students with the same racial or ethnic identity as their student, served a lower proportion of low-income students, had higher proficiency rates on standardized tests, and were more convenient to access (e.g., closer to

home, on a bus line). However, higher income and White caregivers prioritized elementary schools (a) serving students of the same race or ethnicity and (b) fewer low-income students *more* than lower income, Black, and Latine caregivers.¹

Schools that are flagged as needing improvement are not only faced with losing students, but also lose effective teachers due to their categorization. Schools that are ranked poorly on school quality measures are more likely to lose teachers (Feng et al., 2018; Ingersoll et al., 2016). Further, Feng and colleagues (2018) found that the teachers that choose to leave are more likely to be considered “high quality” teachers. Unsurprisingly, the emphasis school accountability systems place on standardized testing plays an important role in retaining teachers, especially in schools facing additional pressures. For example, von der Embse and colleagues (2016) found that increased test-based accountability pressure—as measured by teacher perceptions of the importance of test scores for their own evaluations—was associated with increased teacher stress. High quality teachers are a critical resource for schools. Ideally, under ESSA, schools that are not meeting performance benchmarks should be receiving additional supports and resources. Additional research is warranted on the extent to which school accountability is related to the loss (or gain) of school resources, and the subsequent impact on student outcomes.

Ecological Perspectives on School Accountability

School accountability and improvement systems can play an important role in supporting schools and districts to identify and meet student needs (Cardichon & Darling-Hammond, 2017). However, there are several limitations to current school accountability practices, including limitations to the validity of school quality measures and the potential consequences for schools that are penalized under these practices. An additional limitation is that the extent to which

¹ I use *Latine* as a gender-neutral term for Hispanic and Latina/o populations rather than Latinx (read more here: <https://latv.com/latine-vs-latinx>).

schools can directly impact the metrics used in school accountability systems is unclear. School is only one system—albeit an impactful, important one—within a student’s social ecology. There are many structural barriers and social circumstances unrelated to academic experiences at schools that influence students’ achievement and—therefore—school quality indicators.

Ecological frameworks provide a useful way to conceptualize interplay among the many structural barriers and social circumstances that impact students’ behaviors and academic achievement (Rappaport, 1987; Trickett, 1984). For instance, in his *Ecological Systems Theory*, Bronfenbrenner posited that individuals are embedded within hierarchical social systems that interact with each other—and with the individual—to shape behavior and development (Bronfenbrenner, 1977, 1979). Ecological frameworks are also a useful heuristic for identifying contextual factors that promote or hinder systems change and intervention effectiveness (Foster-Fishman & Behrens, 2007; Peirson et al., 2011; Tseng & Seidman, 2007), such as school improvement efforts.

Researchers have applied this theory to school systems, identifying the specific factors that influence student health, academic achievement, and behavior at each level of the ecological system (see Table 1; Eccles & Roeser, 1999, 2008). When applied to education, ecological frameworks emphasize the importance of systems change—rather than efforts focused on additional ecological levels beyond individual students—in the transformation of schools, districts, or communities (Eccles & Roeser, 2008; Lewallen et al., 2015).

Table 1

Ecological Level(s) of School Systems

Level	Definition
Individual	Student-level factors and behaviors.

Table 1 (cont'd)

Microsystem	Students' social relationships and proximal environments.
Mesosystem	Interactions between agents in a students' microsystems.
Exosystem	The school environment.
Macrosystem	Community, state, and federal contexts.

In the context of school accountability, critics of accountability policies point out that a large proportion of educational outcomes can be explained by demographic characteristics and underlying structural barriers (e.g., macrosystem-level factors), rather than the quality of educational experiences provided by schools (e.g., microsystem- or exosystem-level factors; Schneider et al., 2021). The extent to which accountability systems are able to account for these factors is limited by the data collected by and available to education systems. Below, I detail relevant structural factors—in other words, factors that represent macrosystem-level dynamics a school is unable to directly control—for school accountability and highlight opportunities to utilize existing administrative data to account for them.

Demographic Characteristics and Educational Outcomes

There is a strong association between socioeconomic status and student achievement outcomes (Harwell et al., 2017; Lacour & Tissington, 2011; Sirin, 2005). Evidence suggests that the gap between low income and high income students' achievement has increased over the last 50 years (Reardon, 2011) fueled by the alarming growth of income inequality and economic segregation of schools in the US (Burdick-Will et al., 2011; Duncan & Murnane, 2014; Snellman et al., 2015). Socioeconomic status, which encompasses a variety of domains such as income, family structure, class, and caregiver education, positively predicts student performance on assessments (Davis-Kean, 2005; Reardon, 2011), graduation rates (Bailey & Dynarski, 2011; Duncan & Murnane, 2014), and access to enrichment activities (Duncan & Murnane, 2014;

Snellman et al., 2015). Students of Color are more likely to attend schools with higher poverty levels (Hussar et al., 2020), but even when accounting for socioeconomic status, racial and ethnic minority youth tend to receive lower assessment scores than White youth (Burchinal et al., 2011; Davis-Kean, 2005). Further, Black, Latine, and American Indian/Alaskan Native students have consistently received lower scale scores on math and reading assessments for the last thirty years, despite all students' average scores growing at similar rates (Hussar et al., 2020).

These patterns are also apparent at the classroom, school, and district levels. The socioeconomic status of a student's peers positively predicts their own academic achievement (van Ewijk & Sleegers, 2010). Schools that serve a higher proportion of students that qualify for free-and-reduced lunch have lower assessment scores (Hegedus, 2018; Hussar et al., 2020). Perry and McConney (2010) found that the average socioeconomic status of a school positively predicts student assessment scores even when accounting for individual students' socioeconomic status (Perry & McConney, 2010). Districts in communities with lower average socioeconomic status have lower math and reading assessment scores, and within each district there are large disparities between Black and Latine students' scores compared to their White peers, some of which—but not all—is explained by differences in socioeconomic status (Reardon, 2016).

Given that schools who serve historically marginalized students tend to be rated as having lower achievement, it makes sense that these schools might be more likely to face the consequences of accountability policies. For example, school closures disproportionately impact communities that have high concentrations of students experiencing economic disadvantage or students of Color (Lee & Lubienski, 2017; Tieken & Auldridge-Reveles, 2019). Additionally, a study of the impact of NCLB requirements demonstrated that economically disadvantaged schools that served Black and Latine students were more likely to be flagged as not meeting

Adequate Yearly Progress *and* requiring federal sanctions, despite similar growth in learning over time as their counterpart schools serving majority White and non-economically disadvantaged students (Kim & Sunderman, 2005).

The differences in educational outcomes seen between students based on demographic characteristics are often referred to as “achievement gaps”. The problem with a sole focus on “achievement gaps” is it can detract from the structural factors that lead to differences in educational outcomes; focusing on students’ academic achievement reinforces negative stereotypes about historically marginalized students and limits innovation in our ability to address these inequities (Chambers, 2009; Gouvea, 2021; Milner IV, 2013). Similarly, due to the structure of accountability systems, schools that serve historically marginalized students are being held accountable to these structural factors. In fact, in Downey and colleague’s (2008) study of student growth on standardized test scores, they found that accounting for summer growth rate—to account for non-school factors—increased growth rates substantially, which they coined “impact” rates, i.e., what a school can reasonably impact.

Researchers have documented a multitude of barriers students face for educational and academic achievement, many of which are outside of a school’s locus of control. For example, students describe poverty, frequent moves, competing obligations (e.g., familial obligations, work), and a lack of social support as some of the reasons why their journey to achieve education is difficult (Drotos et al., 2016; Kenny et al., 2007). Families with limited access to financial resources might not be able to afford materials and experiences that support a child’s academic development, such as extracurricular activities, books, art supplies, trips to museums, etc. (Lacour & Tissington, 2011; Snellman et al., 2015). Further, students in poverty are more susceptible to health issues as a result of their circumstances (e.g., food insecurity leading to

malnutrition or anemia; unsafe housing leading to lead, mold, or pollution exposure), which in turn greatly impacts their ability to attend and engage in school (Pascoe et al., 2016).

Some of these barriers are geographically concentrated, negatively impacting the average achievement levels of schools located in under-resourced or underserved communities (Burdick-Will et al., 2011). For example, schools in neighborhoods with higher levels of blight (e.g., broken windows, vacant homes) and lower levels of social capital have worse academic performance (Garnett, 2014; Smart et al., 2021). Many struggling schools and districts in the US are economically and racially segregated, meaning these barriers disproportionately impact the schools and districts that serve low-income, Black, and Latine students (Burdick-Will et al., 2011; Owens et al., 2016; Reardon, 2016; Snellman et al., 2015). Accounting for student characteristics at the school level might serve as a proxy for structural (e.g., macrosystem-level) barriers that certain student populations are more likely to experience.

School and District Resources

Resources at the school level may have important implications for student outcomes. For instance, higher school staff-to-student ratios (i.e., fewer school staff members per student) are associated with higher rates of bullying and dropping out of high school (Christle et al., 2007; Waasdorp et al., 2011) and lower school counselor-to-student ratios (i.e., more school counselors per student) are associated with lower student disciplinary rates and higher graduation rates (Lapan et al., 2012). Further, students of Color and low-income districts might particularly benefit academically from additional school staff, such as teaching assistants, compared to White students and more affluent districts (Hemelt et al., 2021).

However, although the school environment can mitigate the impact of some of these barriers to education (Lacour & Tissington, 2011), schools are only one system in a student's

social ecology and are themselves embedded within a broader context (Eccles & Roeser, 1999, 2008). School and district level characteristics related to funding, staffing, and enrollment likely reflect structural factors such as poverty, neighborhood segregation, and material deprivation, which are clearly linked to student achievement and unlikely to be shifted from school improvement efforts alone. It is critical to determine the extent to which school accountability policies are holding local education agencies accountable to factors outside of the quality of educational experiences they provide. Examining the relationships between existing data that reflect school and district resources and school quality measures is one step towards understanding the extent to which school performance and consequences under current accountability systems are explained by structural factors.

In this project I capitalized on opportunities to utilize existing educational administrative data to capture the impact of structural factors on student outcomes for districts rated poorly by accountability metrics and a matched comparison sample. By using existing and publicly available data, I was only able to identify “proxy” variables of structural factors (e.g., a high proportion of students who qualify for free or reduced lunch serving as a proxy for a context of economic disadvantage).

THE MICHIGAN CONTEXT

Given the autonomy and administrative authority ESSA grants to states to establish their own school accountability systems (McGuinn, 2016), it is worthwhile to consider the unique educational context of the State of Michigan, the location of the proposed study.

School Accountability in Michigan

Michigan's school accountability plan under ESSA was developed by the Michigan Department of Education (MDE) and approved in November of 2017 (MDE, 2017, 2018a). In line with the transition from NCLB to ESSA, Michigan's new accountability system shifts its focus from punitive actions (e.g., financial sanctions) for schools not meeting performance benchmarks to promoting transparency and providing supports to schools that are not meeting their goals (MDE, 2019a). A key element to the new plan is the School Index System. The School Index system generates school quality scores for each school to identify schools and districts that need additional supports to achieve their performance goals.

School Quality Index Scores

As a part of the School Index System, MDE calculates a School Quality Index (SQI) score following ESSA guidelines for every public school in the state. SQI scores reflect weighted averages of a school's performance on the six different components outlined in ESSA (MDE, 2017). Specifically, MDE sets target benchmarks—representing the state's long-term educational goals—for each component and calculates the percentage of the target benchmark that is met by the school. For the most part, these target benchmarks are set at the 75th percentile of current (as of 2017) statewide performance distributions. To promote equity, for each component the target benchmark applies to both the overall student population and student subgroups.² Student

² Per MDE's (2017) ESSA Plan, subgroups include American Indian or Alaska Native students; Asian students; Black or African American students; Hispanic or Latine students; Native Hawaiian or Pacific Islander students;

subgroups must include at least 30 students, otherwise they are not reported or incorporated into that school's SQI score. Student subgroup performance and the whole student population's performance are weighted equally. To do so, MDE calculates SQI scores for the total school population and each valid student subgroup and averages them together to represent a school's overall SQI score.

As an example, MDE sets the following graduation rate goals for all students and student subgroups: 94.40% graduate within four years, 96.49% graduate within five years, and 97% graduate within six years (MDE, 2017). MDE uses a weighted average of these three graduation rates to calculate the target benchmark for the graduation rate component: four-year rates at 50%, five-year rates at 30%, and six-year rates at 20% (i.e., $[94.40 \times .50] + [96.49 \times .30] + [97.00 \times .20] = 95.547$). MDE calculates each school's average graduation rate using the same weights. MDE estimates the school's progress towards the benchmark by dividing its weighted average graduation rate by the target benchmark (i.e., $\text{school's rate} / 95.547$). That number is then weighted at 10% for the school's overall SQI score.

Some components will not apply to every school. For these schools, the percent of the SQI score accounted for by the missing component is added to the other components proportionally based on their weight in the system (MDE, 2017). For example, if the graduation rate component does not apply to a school (e.g., if it is an elementary school), the new SQI score weights are as follows: Student Growth (37.78%), Student Proficiency (32.22%), School Quality/Student Success (15.56%), English Learner Progress (11.11%), and Assessment Participation (2.22%).

Students of Two or More Races; White students; Economically Disadvantaged students; English Learner students; and Students with Disabilities.

Tiered Support Systems

MDE uses these SQI scores to identify public schools and Public School Academies (i.e., public charter schools) that receive Title I funds that need support using the incremental approach prescribed by ESSA (McGuinn, 2016). Every year, MDE flags schools as needing *Additional Targeted Support*—requiring the least intensive support—if they have at least one subgroup with SQI scores in the bottom five percent of all Title I schools’ SQI scores. Every three years, MDE flags schools as needing *Targeted Support* if any of their subgroup SQI scores are in the bottom five percent of all SQI scores and schools as needing *Comprehensive Support*—requiring the most intensive support—for three reasons: (1) their overall SQI scores fall into the bottom five percent of the state, (2) their subgroup SQI scores have been in the bottom five percent of the state for four years, or (3) they have an average graduation rate below 67%. Districts support these schools to establish and implement a targeted support plan to improve the performance of these students.

Based on these assessments, every three years MDE also identifies schools flagged as needing *Comprehensive Support* and their Local Education Agencies (i.e., public school districts) that need additional supports. In these cases, MDE extends an invitation to the district—or charter school—to enter into a District Partnership Agreement for the school(s). These agreements provide additional supports, with the expectation that schools need to demonstrate increased student achievement within three years in response (Strunk et al., 2019). Specifically, the agreement requires schools and districts to conduct needs assessments, develop support plans, implement an evidence-based school reform strategy, and set goals to be met by 18 months and three years (MDE, 2017). As a form of support, MDE assigns an Implementation Facilitator to provide technical assistance with conducting needs assessments, developing and

implementing school/district improvement plans, identifying and implementing school reform strategies, and improving instruction quality.

Considerations for Michigan School Accountability

Although Michigan’s ESSA plan promotes increased transparency and support for schools that are not meeting their performance benchmarks, the School Index System heavily prioritizes standardized test score proficiency and growth, with little-to-no weight given to non-academic outcomes such as access to educational opportunities or disciplinary practices. The limitations of the US’s school accountability system—the validity of school quality measures, potential consequences of accountability policies, and ecological considerations for student and school barriers—all apply to Michigan. Below, I briefly highlight some considerations specific to the Michigan context.

SQI Scores

Most of the SQI score is generated from standardized test data. The ELA and math portions of three standardized assessments are included in the Student Proficiency and Student Growth components: (1) Michigan Student Test of Educational Progress (M-STEP) scores for third through seventh graders; (2) MI-Access test scores, an alternative to the M-STEP administered to students with disabilities as needed; and (3) Preliminary Scholastic Aptitude Test (PSAT) scores for eighth grade students (MDE, 2017, 2021). MDE asserts that these assessments are well established with a large body of evidence supporting their use (MDE, 2017). However, concerns about the reliability and validity of standardized tests still apply. The US Department of Education (2020) conducted a peer review of Michigan’s assessments and determined that additional evidence was needed for the M-STEP and PSAT. For example, MDE removed several written response items from the M-STEP assessments to reduce the testing burden for students,

and this may limit its ability to assess Michigan's ELA and math learning criteria.

Additionally—although there is little published research on Michigan testing—there is at least one paper where the author critiques the extent to which the ELA portion of the M-STEP is a valid measure of reading and language comprehension, in part due to the copious number of criteria it attempts to assess (Sprouse, 2017).

The Student Proficiency component is calculated based on proficiency rates for these tests, rather than scores. When states determine a cut score for proficiency, they determine the size of achievement gaps (Dahlin & Cronin, 2010). There might still be a large difference between student test scores based on different characteristics that could be hidden with a low proficiency cut score or exaggerated with a high proficiency cut score. In fact, Dahlin and Cronin (2010) examined a sample of elementary student's test scores from 36 schools using proficiency cut scores from 28 different states and found that the achievement gap between low-income and non-low-income youth varied in size (and sometimes disappeared) depending on the state's standards. Examining test score distributions and student growth might provide a more accurate assessment of student performance.

The one component of the index system that includes non-academic outcomes is the School Quality/Student Success component, weighted at 14% of the SQI. School Quality/Student Success is a composite measure of the prevalence of chronic absenteeism (K-12 schools), access to advanced coursework (high schools, for grades 11-12), enrollment in postsecondary education (high schools), access to arts/physical education (K-8 schools), and access to librarians or media specialists (K-8 schools) at the school (MDE, 2017). These indicators are weighted equally. Although it is exciting that MDE expanded the definition of what schools do beyond student achievement, the manner in which these indicators are grouped and weighted suggests they are

the least important. MDE frames the new approach to Michigan school accountability under ESSA as shifting its emphasis to a more holistic education and promoting equity (MDE, 2017). However, the SQI score still represents a limited perspective of what schools provide, lacking indicators that address more holistic educational outcomes such as socioemotional learning, access to resources (e.g., school nurses, counselors), or discipline.

Potential Consequences of Michigan Accountability Systems

MDE's ESSA plan outlines when schools and districts are considered to be in violation of accountability policies but does not describe the specific consequences (MDE, 2017), whereas the Partnership Agreements do. Schools and districts that fall under the purview of Partnership Agreements have specific benchmarks that need to be met within 18 months and three years of the agreement being signed (Strunk et al., 2019) and the consequences of not meeting these benchmarks (described as a breach of the plan) can be extreme. For example, most Partnership Agreements include extreme consequences for schools that are in breach of the plan—although they vary on when extreme intervention is required (i.e., at the 18-month or 36-month mark)—such as replacing all school staff, having an Intermediate School District³ or the state take control, or closing the school.⁴

School accountability in Michigan is made more complex by school choice policies, which have been in place since the 1990's and enable parents and caregivers to easily send students to a school or district of their choice (Arsen et al., 1999). The intentions of such policies are to incentivize schools to increase their quality of education and give families autonomy over their child's education (Arsen et al., 1999). However, schools that lose students under these

³ An Intermediate School District is a county- or community-level educational agency that oversees multiple local education agencies (i.e., public school districts).

⁴ Partnership Agreements are publicly available at https://www.michigan.gov/mde/0,4615,7-140-81376_79956---.00.html

policies are not always in a position to increase resources (Arsen & Ni, 2012). Consequently, schools that are subject to district or state intervention under accountability policies could lose more students due to school of choice and not have the resources to bring them back. Enrollment drives school funding, and the costs of educating students facing additional barriers (e.g., students with disabilities) may not be adequately accounted for in Michigan's current funding model (Arsen et al., 2019), exacerbating the impact of declining enrollment in struggling schools.

Ecological Considerations for Students and Districts in Michigan

In Michigan, students who are Black, Latine, qualify for free-or-reduced lunch, or have an Individualized Education Plan are categorized as not proficient in English Language Arts and Math at higher rates, as chronically absent at higher rates, and as not graduating on track at lower rates, compared to state averages (CEPI, n.d.). This means that schools and districts that serve these students are more likely to not meet the performance benchmarks outlined in MDE's accountability system. Many of these inequities are impacted by structural factors. For example, Black and Latine high school students in Michigan report feeling unsafe getting to school, having asthma, and not having access to healthcare at higher rates than their White counterparts (CDC, 2019).

Despite facing more costs in educating the students they serve, districts in Michigan that serve students who are impacted by achievement gaps (e.g., African American students, students with disabilities) are more likely to have lower student achievement, experience declining enrollment, and be in financial distress (Arsen et al., 2016). Further, the amount of money a district has (i.e., fund balance) is positively associated with the number of teachers that are rated as highly effective or effective (Lenhoff et al., 2018). Michigan districts that serve historically

marginalized students are more likely to be impacted by school accountability policies, but less likely to have the resources to address the barriers their students face.

THE CURRENT PROJECT

As states submitted their ESSA plans for approval in 2017, it is an opportune time to contribute to the burgeoning body of research examining potential implications of these new accountability policies. Educational agencies and legislators position ESSA as a new wave of accountability that promotes high-quality, equitable, and well-rounded education (Adler-Greene, 2019; Darling-Hammond et al., 2016). However, many researchers have raised concerns about school accountability policies, particularly regarding the validity of the measures used to flag schools as needing improvement and the potential consequences for schools that are flagged as such (Bross et al., 2016; Schneider et al., 2021; Sunderman et al., 2017). An ecological perspective of school accountability highlights the structural barriers that impact student performance, many of which are outside of what a school can and should be expected to impact (Eccles & Roeser, 1999, 2008), and that the schools most likely to be impacted by school accountability policies are often situated within contexts that limit their ability to address them.

My project builds upon these considerations and addresses multiple areas of research that merit expansion. I conducted an exploratory study to examine how schools in districts with Partnership Agreements are impacted by structural factors—measured by proxy variables generated from publicly available administrative data—and determine the degree to which these factors account for student outcomes, relative to a matched comparison sample of districts rated as better performing. I was able to generate three categories of measured indicators of structural factors (e.g., proxies for macrosystem-level factors) based on available data: (1) student enrollment characteristics (e.g., demographics); (2) student mobility characteristics (e.g., mobility rate); and (3) school and district resources (e.g., staff-to-student ratios).

In this project, I also included discipline as a student outcome of interest, and access to financial and personnel resources as predictors of student outcomes. Few states meaningfully incorporate non-academic indicators in their school quality measures (Dusenbury et al., 2018; Hough et al., 2017; Schneider et al., 2021), leading to a lack of research on the extent to which exclusionary discipline practices or school resources are related to school quality.

I also compared the extent to which these schools are achieving student outcomes equitably across major student subgroups. ESSA heavily emphasizes equity by way of addressing documented achievement gaps among historically marginalized students. Additional research is needed to determine whether equitable outcomes are associated with school quality measures. Researchers have identified competing perspectives of this topic: (1) higher performing schools have increased resources and are therefore better able to serve all students, including those historically disenfranchised by the education system or (2) higher performing schools tend to serve populations that are not historically disenfranchised by the education system, but the students that are do not fare any better (and perhaps, fare worse) (Chambers et al., 2014; Gaddis & Lauen, 2014; Harris & Herrington, 2006).

Research Questions and Hypotheses

I examined longitudinal trends in schools in districts that had entered Partnership Agreements with MDE as of September 2021 relative to a matched comparison sample. In Michigan, the District Partnership Agreements are a path to increasing resources in districts that are categorized as underperforming, but these agreements operate under the assumption that districts can address student achievement with increased resources. My primary aims were to examine the extent to which: (1) partnership district schools differ from matched comparison district schools on student outcomes over time; (2) proxies for structural factors (e.g., enrollment,

financial status) impact resources and student outcomes; and (3) partnership district schools differ from matched comparison district schools in terms of equity.

My research questions and hypotheses were:

RQ1: How do partnership district schools—relative to their matched comparisons—experience changes in student outcomes over time?

H1: *Partnership district schools will experience more declines in student outcomes relative to their matched comparison over time.*

RQ2: How do proxies for structural factors predict student outcomes?

H2: *Indicators of structural barriers will predict worse student outcomes; indicators of structural resources will predict better student outcomes.*

RQ3: Are the patterns identified in RQ1 and RQ2 the same when examining outcomes for different groups of students?

H3: *Will clarify competing hypotheses: (1) Matched comparison district schools have more equitable student outcomes (i.e., student outcome trends are positive and similar across subgroups), and these outcomes are predicted by structural factors; or (2) Partnership and matched comparison district schools do not differ in equitable student outcomes (i.e., student outcome trends are similar across subgroups for partnership and matched comparison district schools), and these outcomes are not predicted by structural factors.*

METHODS

Data Sources

I used two sources of publicly available education data from the 2009-10 school year to the 2018-19 school year: MI School Data and Civil Rights Data Collection. Even though Michigan recently implemented the current accountability system, some of the data is available going back to 2009-10. I selected this timeframe to increase my ability to detect longitudinal trends. I did not include data post 2018-19 due to the onset of the Coronavirus pandemic in the 2019-20 school year and its tremendous impact on students, schools, and the student data available for that year.

Although there are many benefits to using administrative data, one challenge with my project was that there was an abundance of data yet limited guidance on finding or defining reliable and valid indicators for my constructs of interest. Thus, I employed an exploratory approach where I included many different variables—some seemingly redundant—to empirically derive which indicators have predictive validity, given my research questions.

MI School Data

MI School Data (<https://www.mischooldata.org/>) is a data hub where the State of Michigan houses public education data. MI School Data is operated by Michigan's Center for Educational Performance and Information (CEPI) and is the primary source of data reported from all Michigan public schools and districts to the Department of Education. Data include information on finances, staffing, educational opportunities, school quality, student achievement, and post-secondary pathways. Data are reported at the student subgroup (e.g., by grade), school, and/or district levels every year. There are instances where the number of individuals in any given category is small enough to be identifiable and CEPI omits these data points to protect the

privacy of students and schools. Specifically, when fewer than 10 students are in a category their data is not reported, any values less than 5% are coded as 5%, and any values greater than 95% are coded as 95% (MDE, 2017).

Civil Rights Data Collection

Civil Rights Data Collection (CRDC; <https://ocrdata.ed.gov/resources/downloaddatafile>) is a data hub where the US Department of Education houses public education data. CRDC is operated by the Office for Civil Rights and collects and reports information related to school practices and services broken down by different student characteristics to monitor and ensure adherence to civil rights statutes. School- and district-level data are reported every other year for each state. It is important to note that the CRDC website indicates many of the outcomes are underreported, which limited my ability to include certain variables of interest, such as the prevalence of harassment and bullying.

Sampling

For my project, I focused on all the schools within Local Education Agencies (i.e., public school districts) in Michigan that entered Partnership Agreements (n = 12) as of September 2021 and identified schools from higher performing districts (n = 12) to serve as a matched comparison sample. I did not include charter schools that have entered Partnership Agreements.

Matching Procedure

To identify the comparison sample, I matched the districts with Partnership Agreements (i.e., partnership districts) one at a time using the procedures outlined in Table 2. As a first step, I randomized the order in which I matched the partnership districts with their comparison district, as multiple districts might have the same district emerge as the best match. In this order, I

matched the partnership districts one at a time based on community, school, and student demographic characteristics (outlined in Steps 2 and 3 in Table 2 below).

I used data from the 2009-10 school year to match districts based on their characteristics *at baseline*, due to my hypothesis that declining enrollment over time will be a defining characteristic of partnership districts relative to their matched counterparts. For these characteristics, I prioritized matching based on community level factors, economic disadvantage, and district size. I anticipated systematic differences between the demographics of the partnership districts and their matched comparisons, in particular the proportion of students that are identified as economically disadvantaged, qualifying for special education services, and Black or African American. Upon initial exploration I found that matching on all three characteristics was not possible, and that economic disadvantage was (a) the most feasible and (b) would allow me to look at the roles of race and special education more clearly. As a final step, I calculated the average School Quality Index score for the partnership districts and the potential matched districts to ensure that—despite similar characteristics—the matched district was relatively higher performing per this index.

Table 2

Matching Procedures

1. Randomly assigned each partnership district a number and sorted them lowest to highest to determine the order of matching.
2. Identified the districts that matched the partnership district at baseline (2009-10) based on the following characteristics: a. Locale (urban, rural, suburban; small, mid-size, large) b. School level (e.g., elementary through high school) c. Student counts (enrollment) i. I set a threshold of $\pm 25\%$ of the number of students in the partnership district.
3. Removed any districts that were not open as of the 2018-19 school year.

Table 2 (cont'd)

4. Sorted through the remaining districts to determine which one had the most similar student demographics at baseline (in the following order):
a. Economically disadvantaged
b. Student counts (enrollment)

5. Calculated the average School Quality Index (SQI) scores (scored from 0-100) for 2017-18 and 2018-19 for the potential matched comparison and determined whether it was sufficiently higher than the corresponding partnership district by:
a. Calculating the average SQI scores for 2017-18 and 2018-19 for the partnership districts:
i. Calculated the mean index score for each district's general and special education schools, excluding alternative schools which included adult and vocational programs.
ii. Aggregated these mean scores by year to create an overall partnership district mean SQI.
b. Calculating the standard deviations of the partnership districts' mean SQI scores for 2017-18 and 2018-19.
c. Calculating the potential matched comparison's mean SQI score:
i. Calculated the mean index score for the district's general and special education schools, excluding alternative schools which included adult and vocational programs.
ii. Aggregated these mean scores by year to create an overall matched comparison district mean SQI.
d. Determined whether the matched comparison's mean SQI score was at least one standard deviation higher than the partnership district's for both 2017-18 and 2018-19.
i. If it did not meet this standard, I returned to Step 2 and repeated the process until I found the most similar district with a sufficient index score.

I had to repeat the process due to an insufficient School Quality Index score of the first potential match for four partnership districts. Two were resolved by selecting the second- or third-closest match available after filtering based on the locale, school level, and enrollment thresholds established in Step 2. For the other two districts, the only match available after setting the thresholds in Step 2 and examining the School Quality Index scores was too dissimilar in economic disadvantage (8% compared to 84%, 8% compared to 76%). To address this, I expanded the enrollment range to $\pm 35\%$ of the partnership district's total baseline student count. One of the partnership districts was the only district categorized as being in a "large city" and

had the largest enrollment in the state ($n = 88,218$). There were no districts within $\pm 25\%$ or $\pm 35\%$ of its student count so I selected the next three districts with the highest enrollment ($n = 29,325$, $n = 19,088$, $n = 18,905$) as potential matches. For these three districts, I completed Steps 3 through 5 to identify the matched comparison.

Sample

The resulting sample included the partnership districts and their matched comparisons ($n = 24$), all of which served elementary through high school students. The districts came from 15 counties across Michigan and were predominantly categorized as located in cities ($n = 14$, 58%) or suburbs ($n = 8$, 33%). Only two districts were categorized as rural. I provide a summary of the district-level demographic characteristics of the partnership districts and of the matched comparison districts in Table 3.⁵ On average, partnership districts had higher student counts, higher proportions of Black students and of students categorized as economically disadvantaged, and a lower proportion of White students. When I dropped the partnership district with nearly 90,000 students, average student counts for partnership districts were comparable to their matched comparisons at baseline (Mean [M] = 6982.27, Standard Deviation [SD] = 6184.41).

As of 2018-19 there were 408 schools across these 24 districts (range = [407, 540]). I only included schools that were categorized as serving elementary through high school students, were not categorized as a unique education provider (e.g., alternative programs, adult education), and enrolled more than 30 students.⁶ The resulting sample included 392 schools as of the 2018-

⁵ Any demographic information summarized with means in Table 3 reflects unweighted grand means. I chose to calculate unweighted means for this table because I aimed to describe—on average—who the different districts tend to serve. I did not want the smaller district's demographics to be underrepresented, given that student counts vary substantially within and between district types.

⁶ While exploring the data, I noticed multiple school names with few students (e.g., $n=5$) enrolled that appeared to be an alternative education program but were not categorized as such (e.g., Home Education Site or Adult Education Program). To address this, I added a cutoff for enrollment of 30 students or more.

19 school year (range = [392, 502]). Student demographics appeared similar for both sets of districts between the 2009-10 school year and the 2018-19 school year (Table 3).

Table 3

Summary of District Characteristics and Student Demographics

District Information	Partnership	Matched Comparison
	<u>N (%)</u>	<u>N (%)</u>
Rural	1 (8.30%)	1 (8.30%)
Suburb	4 (33.30%)	4 (33.30%)
City	7 (58.30%)	7 (58.30%)
	<u>Mean (SD)</u>	<u>Mean (SD)</u>
District Average SQI 2017-18	42.75 (9.74)	66.16 (8.41)
District Average SQI 2018-19	41.93 (9.46)	66.47 (9.34)
2009-10 Student Demographics	Partnership	Matched Comparison
	<u>Mean (SD)</u>	<u>Mean (SD)</u>
Student Count	13,752 (24,181)	7,232 (6,203)
Student Gender		
Male	52.16% (1.42%)	52.12% (1.10%)
Female	47.84% (1.42%)	47.88% (1.10%)
Student Race/Ethnicity		
African American/Black	55.42% (20.74%)	14.29% (14.33%)
Caucasian/White	28.36% (19.34%)	71.56% (21.06%)
Hispanic/Latine	10.55% (5.73%)	9.31% (11.70%)
Asian	1.07% (1.19%)	2.23% (2.28%)
Hawaiian/Pacific Islander	0.05% (0.04%)	0.16% (0.13%)
AI/Alaskan Native	0.61% (0.33%)	0.79% (0.41%)
Two or more races	3.96% (7.21%)	1.66% (2.25%)
Student Characteristics		
Economically disadvantaged	77.22% (11.35%)	59.22% (11.16%)
Enrolled in special education	16.66% (3.90%)	12.48% (2.72%)
English Language Learners	9.00% (7.14%)*	10.01% (11.56%)*
2018-19 Student Demographics	Partnership	Matched Comparison
	<u>Mean (SD)</u>	<u>Mean (SD)</u>
Student Count	8,656 (13,804)	6,693 (6,262)
Student Gender		
Male	52.60% (1.63%)	51.60% (0.68%)

Table 3 (cont'd)

Female	47.40% (1.63%)	48.40% (0.68%)
Student Race/Ethnicity		
African American/Black	56.32% (21.61%)	13.75% (11.91%)
Table 3 (cont'd)		
Caucasian/White	21.89% (17.28%)	63.54% (22.46%)
Hispanic/Latine	14.19% (10.65%)	13.31% (14.15%)
Asian	1.19% (1.22%)	2.60% (3.80%)
Hawaiian/Pacific Islander	0.14% (0.27%)	0.07% (0.05%)
AI/Alaskan Native	0.30% (0.20%)	0.36% (0.21%)
Two or more races	5.98% (4.86%)	6.36% (3.59%)
Student Characteristics		
Economically disadvantaged	81.27% (10.44%)	66.81% (10.14%)
Enrolled in special education	16.84% (3.21%)	12.89% (2.34%)
English Language Learners	8.75% (8.46%)*	12.89% (14.95%)*

Note. *At times, the number of English language learner students was intentionally omitted to protect student anonymity. These averages are likely overestimated as a result.

To confirm that—on average—matched comparison district schools were categorized as higher performing than partnership district schools, I calculated hedges g to examine the difference between partnership district schools and matched comparison district schools' SQI scores. Hedges g is a measure of effect size that describes the size of a difference between two groups, accounting for small sample sizes (Hedges, 1981). Across both years, partnership district schools had an average SQI score of 39.74 ($SD = 19.68$) and matched comparison district schools had an average SQI score of 63.49 ($SD = 23.11$), and the adjusted difference between these two averages was large ($g = 1.12$, 95% Confidence Interval [CI] = [0.97, 1.28]).

Despite my best efforts to match districts based on student economic status, the proportion of students categorized as economically disadvantaged appeared to be higher at baseline for partnership districts ($M = 77.22\%$) compared to matched comparison districts ($M = 59.22\%$) at the district level. To determine the size of this difference, I calculated hedges g for

the difference of the proportion of economically disadvantaged students at baseline between partnership district schools and matched comparison district schools. In 2009-10, partnership district schools served an average of 79.26% ($SD = 16.09\%$) of economically disadvantaged students and matched comparison district schools served an average of 59.36% ($SD = 19.88\%$). The adjusted difference between these two averages was large ($g = -1.14$, 95% CI = [-1.34, -0.94]).

Variables and Constructs

I examined proxies for structural factors—which include constructs capturing enrollment, mobility, and school and district resources—as predictors of student outcomes, which include attendance, assessment, and disciplinary outcomes. I also accounted for school- and district-level school characteristics. I define each construct and its indicators below. I provide the complete list of variables, their sources, and their measurement frequency in Appendix A.

Measured Indicators of Structural Factors

I sourced measured indicators from my data sources to serve as proxies of structural factors; these measured indicators captured three overarching constructs: enrollment, mobility, and school and district resources. Due to the large number of related indicators, I employed factor analysis to reduce and combine the indicators in the multilevel models. Below, I describe each measured indicator prior to describing the factor analysis approach and results in a subsequent section.

Student Enrollment. Student enrollment includes different elements that summarize the students enrolled in a school or district. I included the number of students enrolled, as well as broken down by demographic characteristics.^{7,8} All indicators were available at the school level.

Overall Enrollment. The total number of students each year from the 2009-10 school year to the 2018-19 school year.

Enrollment by Demographic Background. The proportion of students enrolled broken down by race and ethnicity (Black, White, Hispanic or Latine)⁹ each year from the 2009-10 school year to the 2018-19 school year, as well as the proportion of students that are classified as economically disadvantaged (i.e., qualify for free and reduced lunch [FRL]) and qualify for special education services (i.e., have an Individualized Education Plan [IEP]) each year from the 2009-10 school year to the 2018-19 school year.

Mobility. Student mobility includes different elements that summarize the movement of students in a school or district. I examined the proportion of students who utilized school of choice and the mobility rate of all students as well as broken down by demographic characteristics. Most indicators were available at the school level with the exception of school of choice enrollment, which was only available at the district level.

Mobility Rate. The ratio of the number of students who are classified as mobile (i.e., leave the district for whatever reason) to the total number of students each year from the 2009-10 school year to the 2018-19 school year.¹⁰

⁷ I used the same threshold as ESSA for subgroup inclusion and focus on outcomes for the groups with school-level averages > 30. For example, there are 7 racial and ethnic categories provided in the data, however, I will only examine outcomes for Black, White, and Latine student subgroups.

⁸ Due to the high number of instances schools omitted the number of English Language Learner students and their outcomes (N = 1959, 45.7%), I was not able to examine this subgroup of students.

⁹ Different reports provide slightly different demographic categories (e.g., African American vs Black). I assumed these categories were equivalent—as they come from the same data—and used the category names I felt were most inclusive.

¹⁰ As of 2012-13, students who graduated on or after April 25 were not categorized as mobile students.

Mobile Student Demographics. The proportion of mobile students broken down by race and ethnicity (Black, White, Hispanic or Latine) each year from the 2009-10 school year to the 2018-19 school year, as well as the proportion of students that are classified as economically disadvantaged (i.e., qualify for FRL) and qualify for special education services (i.e., have an IEP) each year from the 2009-10 school year to the 2018-19 school year.

School of Choice Enrollment. The ratio of the number of students that live within district boundaries but attend another district to the total number of students each year from the 2011-12 school year to the 2018-19 school year.

School and District Resources. School and district resources include different elements that indicate the fiscal and personnel resources available to a school or district. Specifically, I included school and support staff information, which were available at the school level, as well as financial resources, which were only available at the district level. Support staff data was only available every other year.

School Staff. Academic staff includes the ratio of instructional paraeducators, non-instructional paraeducators, and teachers to the total number of students each year from the 2009-10 school year to the 2018-19 school year.

Support Staff. Support staff includes the ratio of school counselors, nurses, social workers, and psychologists to the total number of students over time every other year. The number of school counselors to students was available from the 2009-10 school year to the 2017-18 school year. The number of nurses, social workers, and psychologists was available from the 2015-16 school year to the 2017-18 school year.

Financial Characteristics. Financial characteristics include the ratio of the district's total revenue (in dollars), fund balance (in dollars), and total expenditures (in dollars) to the total number of students each year from the 2011-12 school year to the 2018-19 school year.

Student Outcomes

Student outcomes include attendance, assessment, expulsions, and suspensions for the overall school population and for each of the student subgroups, all of which were available at the school level. Suspension data were only available every other year. Expulsion data were not provided at the subgroup level.

Attendance. Attendance includes the overall average attendance rate of students (the number of days present divided by the total number of days enrolled).¹¹ I included attendance rates for the overall student population as well as broken down by race and ethnicity (Black, White, Hispanic or Latine), economic status (qualify for FRL), and special education services status (have an IEP) each year from the 2011-12 school year to the 2018-19 school year.

Assessment. Assessment outcomes include average student scores and growth percentiles for students grades three through eight on the English Language Arts (ELA) and math domains of the M-STEP assessment each year from the 2014-15 school year (for scores) and 2015-16 school year (for growth percentiles) to the 2018-19 school year.¹² These outcomes are included for the overall student population as well as broken down by race and ethnicity (Black, White, Hispanic or Latine), economic status (qualify for FRL), and special education services status (have an IEP).

¹¹ Prior to 2017-18, students were marked absent when they missed the entire day. Beginning in 2017-18, students are marked absent when missing 50% or more of the day.

¹² As of Spring 2019 (which is outside the study timeline), eighth graders only take the M-STEP social studies assessment, instead taking the PSAT for math and ELA (MDE, 2019b).

The goal of the M-STEP is to assess whether students are meeting their grade-level standard (MDE, 2019b). M-STEP items come in three forms: (1) multiple choice items (which make up the majority of the test), where students select one correct answer from multiple options; (2) technology enhanced items, which are more interactive than multiple choice and might ask students to respond to a question by highlighting specific words in a paragraph or matching an image to a passage; and (3) short answer questions, where students type a passage or essay in response to a prompt (MDE, 2018b). MDE partners with the Smarter Balanced Assessment Consortium (Smarter Balanced) and the Data Recognition Corporation to develop and score the M-STEP ELA and math assessments (MDE, 2019b).

Assessment Scores. Assessment scores include the mean ELA and math M-STEP scale scores for each school. Scale scores are standardized scores that adjust for different forms of the same test (e.g., if Form A contains more items than Form B, the raw scores need to be adjusted so that scores are equivalent across forms). MDE transforms M-STEP raw scores into scale scores—with specific cut scores that reflect a student’s proficiency—within each grade and content area using psychometric procedures established by Smarter Balanced (Smarter Balanced, 2018). Briefly, Smarter Balanced constructed scales for each M-STEP test using Item Response Theory models during the pilot stages of test development. Smarter Balanced maps individual M-STEP items onto these scales by determining the knowledge required to respond and the response probability, or the likelihood of a student providing the correct answer (MDE, 2018b; Smarter Balanced, 2018).

Because these scaling procedures are completed for each grade’s assessment, the range of M-STEP scaled scores varies between content and grade level. Different ranges indicate different proficiencies, which are described by the following “performance levels”: Not Proficient,

Partially Proficient, Proficient, and Advanced (see Appendix B for a breakdown of the performance levels and their corresponding score ranges as of the 2018 administration of the M-STEP). To address the fact that these scores were only comparable within grades and content areas, I calculated z scores for each grade level and each assessment and aggregate within content areas (resulting in an average z score for each school's ELA and math scores). I multiplied these scores by 10 for analysis to facilitate interpretation of the parameter estimates.

Assessment Growth Percentiles. Assessment growth includes the mean ELA and math growth percentiles. A growth percentile captures a student's growth in their test scores from the previous year relative to students in the same grade who had similar test scores in the same content area (MDE, 2019b). Growth percentiles are calculated across all students in the state but are only available for students with valid scores on the current and previous assessments. Growth percentiles range from 0-100, and different ranges indicate different proficiencies: Below Average Growth (1st-29th percentiles), Average Growth (30th-69th percentiles), and Above Average Growth (70th-99th percentiles).

Expulsions. Expulsions include the ratio of the number of students who received expulsions to the total number of students each year from the 2009-10 school year to the 2018-19 school year. On MI School Data, Expulsions data only includes a list of schools that report one or more expulsions each year. As such, I assigned any schools that were missing data a value of 0. There is a chance that a school expelled students but did not report them.

Suspensions. Suspensions include the proportion of students who received in-school suspensions (the number of students who received one or more in-school suspensions to the total number of students) and the proportion of students who received out-of-school suspensions (the number of students who received one or more out-of-school suspensions to the total number of

students) every other year from the 2009-10 school year to the 2017-18 school year. Suspension outcomes are included for the overall student population as well as broken down by race and ethnicity (Black, White, Hispanic or Latine) and special education service status. Suspension data come from CRDC, which uses two methods to identify special education students: students served under the Individuals with Disabilities Education Act (IDEA) and students served under Section 504 of the Rehabilitation Act (but not served under IDEA). Because all students served under IDEA are required to have an IEP and are also covered under Section 504 (cite <https://sites.ed.gov/idea/>), I used the number of students served under IDEA to reflect students qualifying for special education services. CRDC does not report suspension rates broken down by economic status.

School and District Characteristics

School and district characteristics include school accountability elements I account for in my models.

School Quality Index. As described in the introduction, for the past few years MDE has generated index scores for each school as indicators of school quality per ESSA guidelines. SQI scores range from 0-100 and reflect weighted averages of the following elements (see Figure 1 on pg. 25 for a visual depiction): 34% standardized test score growth, 29% student proficiency on standardized tests, 14% school quality/student success (this includes chronic absenteeism, enrollment in advanced coursework, postsecondary enrollment, and access to arts and physical education), 10% graduation rate, 10% English Language Learner progress; and 3% assessment participation (MDE, 2017).

There are currently no thresholds established for what a “good” School Quality Index score is; the lowest performing schools and districts are identified as those with the lowest scores

relative to the rest of the state. However, MDE (2017) reported results of initial models of SQI scores and determined that 23% of schools have scores from 90-100, 29% have scores from 80-89, 22% have scores from 70-79, 13% have scores from 60-69, and 14% have scores below 60. I calculated an average of each school's index score for the two years they were available: 2017-18 and 2018-19.

District Type. I effect coded the type of district as either “partnership” (1) or “matched comparison” (-1) to allow me to compare findings between the two samples.

Missing Data

There are two major sources of missingness within the current sample: (1) missing MI School Data due to intentional data omission and (2) missing CRDC data due to not reporting. MI School Data omits data for student outcomes that are representative of fewer than 10 students. To understand the extent to which this intentional omission impacted the current sample, I looked at the proportion of missing student data that could be attributed to said omission in each student subgroup (e.g., the number and percent of cases that did not report the attendance rate for students with an IEP that had fewer than 10 students with an IEP enrolled). I accounted for the fact that different variables had different total possible observations. For example, mobility rates were available from 2009-10 to 2018-19 for all schools with 4198 possible observations, whereas M-STEP scores were available from 2014-15 to 2018-19 for all schools except high schools with 1612 possible observations. Data omission accounted for most missing student subgroup data (see Table 4).

Table 4*Missing Data Affected by Omission*

	Student Subgroup				
<u>Omission</u>	Black	White	Latine	FRL	IEP
<u><10 Enrolled</u>	<u>N (%)</u>	<u>N (%)</u>	<u>N (%)</u>	<u>N (%)</u>	<u>N (%)</u>
(%)	942 (22.4%)	1462 (34.8%)	2841 (67.7%)	0 (0%)	294 (7.0%)
<u>Mobility Rate</u>					
Total Missing	155 (3.7%)	328 (7.8%)	527 (12.6%)	139 (3.3%)	1473 (35.1%)
Omitted	32 (20.6%)	291 (88.7%)	490 (93.0%)	0 (0%)	283 (19.2%)
<u>Attendance Rate</u>					
Total Missing	221 (6.8%)	406 (12.5%)	633 (19.6%)	73 (2.3%)	186 (5.7%)
Omitted	139 (62.9%)	391 (96.3%)	623 (98.4%)	0 (0%)	113 (60.8%)
<u>ELA Scores</u>					
Total Missing	536 (33.3%)	712 (44.2%)	1103 (68.4%)	151 (9.4%)	893 (55.4%) ^a
Omitted	345 (64.4%)	474 (66.6%)	954 (86.5%)	0 (0%)	96 (10.8%)
<u>Math Scores</u>					
Total Missing	538 (33.4%)	714 (44.3%)	1100 (68.2%)	151 (9.4%)	892 (55.3%) ^a
Omitted	345 (64.1%)	474 (66.4%)	952 (86.6%)	0 (0%)	96 (10.8%)
<u>ELA Growth</u>					
Total Missing	437 (27.7%)	644 (40.9%)	942 (59.8%)	167 (10.6%)	498 (31.6%)
Omitted	258 (59.0%)	466 (72.4%)	824 (87.5%)	0 (0%)	132 (26.5%)
<u>Math Growth</u>					
Total Missing	436 (27.7%)	642 (40.7%)	945 (60.0%)	167 (10.6%)	500 (31.7%)
Omitted	257 (58.9%)	467 (72.7%)	825 (87.3%)	0 (0%)	132 (6.4%)

- a. Omission accounted for a small proportion of missingness on M-STEP scores for IEP students; Michigan offers an alternative exam for certain students with cognitive disabilities, the scores of which were not in the purview of this study.

In addition to the observations above, there were several MI School Data variables for which missingness would not be affected by omission (e.g., number of teachers, outcomes for all students; see Table 5). Further, although the CRDC survey is required for most public schools, a

number of schools in the current sample did not report disciplinary or school support staff data to CRDC.¹³

Table 5

Missing Data Not Affected by Omission

	Total Missing N (%)
Mobility Rate	130 (3.1%)
Percent Chronically Absent	243 (8.2%)
Chronically Absent attendance rate	243 (8.2%)
Attendance Rate	73 (2.3%)
ELA Scores	147 (9.1%)
Math Scores	147 (9.1%)
ELA Growth	155 (9.8%)
Math Growth	155 (9.8%)
Number of Counselors ^a	396 (18.5%)
Number of Nurses ^b	15 (2.0%)
Number of Psychologists ^b	15 (2.0%)
Number of Social Workers ^b	15 (2.0%)
Out-of-school Suspensions ^a	183 (8.5%)
In-school Suspensions ^a	182 (8.5%)

a. Available every other year, from 2009-10 to 2017-18

b. Available 2015-16 and 2017-18

Outliers

I set thresholds for certain variables to identify implausible values (e.g., an attendance rate higher than 100%) given that my sample focuses on general education schools that serve more than 30 students grades K-12. I provide these thresholds and the number of observations that met them in Table 6. In the case of M-STEP scores, I used the scale options provided in Appendix B (as the range of possible scores differs by grade level). I removed implausible observations as estimates generated from multilevel modeling are robust to randomly missing observations (Hox et al., 2018).

¹³ <https://www2.ed.gov/about/offices/list/ocr/frontpage/faq/crdc.html>

Table 6*Implausible Value Thresholds by Variable*

Construct	Threshold(s)	Cases
Enrollment by demographics	>100%	0
Mobility rates	>100%	0
Attendance rates	>100%	0
Assessment scores	<i>See Appendix B</i>	0
Assessment growth	>100%	0
% In-school suspensions	$\geq 100\%$	8
% Out-of-school suspensions	$\geq 100\%$	86
% Expelled	$\geq 100\%$	0
School staff-to-student ratios	≥ 1	1
Support staff-to-student ratios	≥ 1	2
SQI	>100	0

I then examined univariate outliers by computing z-scores, using a cutoff value of 3.29 (Tabachnick & Fidell, 2018). In a multilevel context, it is best to examine outliers within each level (Langford & Lewis, 1998). As such, I examined outliers across districts, across schools (within districts), and across time (within districts, within schools; Table 7). I examined outliers across districts by calculating z scores for each district's mean relative to the overall district-level means and standard deviations for each continuous variable. I examined outliers across schools by calculating z scores for each school's mean relative to the overall means and standard deviations of all the schools in the same district for each continuous variable. I examined outliers across time by calculating z scores for each school's mean—or district's mean in the case of district-level variables—in a given year relative to the overall means and standard deviations for that school (or district) over time for each continuous variable.

Overall, there were no outliers across time within schools for any of the variables (Table 7). The most school-level outliers were observed in attendance rates, and the most district-level outliers were observed in the proportion of students receiving in-school suspensions.

Table 7*Outliers by Level of Analysis*

Construct	District^a	School^b	Time^c
Total Enrollment	0	12	0
Enrollment by demographics	1	56	0
Mobility rates	1	54	0
Attendance rates	0	104	0
Assessment scores	0	15	0
Assessment growth	1	14	0
% In-school suspensions	3	80	0
% Out-of-school suspensions	1	59	0
% Expelled	0	14	0
School staff-to-student ratios	2	77	0
Support staff-to-student ratios	2	47	0
SQI	0	3	0
District finances	2	-	0
District school choice	0	-	0

a. Number of districts

b. Number of schools

c. Number of observations

Data Preparation

In this exploratory study, I aimed to use the data as reported to MDE and CRDC. As such, I provide information regarding statistical outliers and missing data as context. I did not employ any imputation, transformation, or deletion in my analyses, aside from deleting implausible values that fell outside the expected range of a given variable. I made one exception to this. While examining outliers and the distribution of the data, I noticed two observations that fell just under the implausible data thresholds for school staff-to-student ratios *and* appeared significantly higher than the second-highest observation (counselor-to-student ratio of 99/100, the next largest value was 4.5/100; instructional paraeducator-to-student ratio of 87/100, the next largest value was 45/100). As a sensitivity analysis, I reran any final models that included school staff or counselors as a statistically significant predictor with these outlying values removed and describe any key differences in the results.

Data Reduction using Confirmatory Factor Analysis

I employed Confirmatory Factor Analysis (CFA) to examine whether any of the measured indicators of structural factors could be more parsimoniously grouped, such that the number of parameters in my multilevel models could be reduced. If a construct (as theorized in the variables section above) had greater than three school-level variables measured at the same time points, I created a weighted indicator using standardized factor loadings generated from a CFA (see Table 8). Given that factor loadings can be negative and the variables of interest are already comparable across schools regardless of size (i.e., represent proportions or ratios), I did not generate weighted averages; I multiplied each variable by its standardized factor loading and then summed each of the weighed variables per construct (Hair et al., 2010).

Four constructs met these criteria: (1) enrollment demographics, (2) mobility, (3) school staff, and (4) support staff. To construct these factors, I used one year of data based on the first year the variables were available (e.g., I constructed enrollment demographics using data from 2009-10). In the case of support staff, I excluded the number of school counselors as that variable was available for a longer time period (starting in 2009-10) and I did not want to lose the prior observations of that variable. I requested modification indices to determine whether there were any meaningful covariances between the variables (e.g., the proportion of students that are Black is dependent on the proportion of students that are White).

CFA Model Fit Indices

I used the following four indices and respective critical values for adequate model fit to assess the fit of the CFA models: (1) Chi-Square, $\chi^2 p > .05$ (Hu & Bentler, 1999; Mulaik et al., 1989); (2) Comparative Fit Index (CFI), $CFI > .90$ (Bentler, 1990; Hu & Bentler, 1999); (3) Root Mean Square Error of Approximation (RMSEA), $RMSEA < .08$ (Hu & Bentler, 1999; Steiger,

1998); and (4) Standardized Root Mean Squared Residual (SRMR), $SRMR < 0.08$ (Hu & Bentler, 1999).

CFA Results

I present the results CFAs examining the fit of my suggested approaches to measuring enrollment demographics, mobility demographics, school staff, and support staff here (see Table 8), as I conducted these analyses as a preliminary step before estimating the models to answer my research questions. I estimated a one-factor structure separately per construct because I aimed to generate weighted indicators of conceptually distinct variables. Even if the constructs were related to each other, their covariance was accounted for in the subsequent models. The practical implication of this approach is that the results allowed me to reduce the number of parameters in my longitudinal models. However, the weights I established from the CFA models are not meant to be applied in future research or to other data; they are all data-dependent.

Table 8

CFA Results

Model Fit						
	N	Parameters	Chi-Square	RMSEA	CFI	SRMR
Enrollment Demographics ^a	502	18	$\chi^2(2) = 12.15^*$	0.10	0.99	0.03
Enrollment Demographics ^b	502	15	$\chi^2(5) = 28.72^{***}$	0.10	0.99	0.04
Mobility Rate	500	18	$\chi^2(5) = 93.18^{***}$	0.14	0.97	0.03
School Staff ^a	499	9	-	-	-	-
School Staff ^b	499	8	$\chi^2(1) = 0.16$	<.001	1.00	<0.01
Support Staff	381	9	-	-	-	-
Parameter Estimates						
	Unstandardized		Standardized			
	<u>B</u>	<u>SE</u>	<u>β</u>	<u>SE</u>		
<u>Enrollment Demographics^a</u>						
1. % of Black students	32.20	17.96	0.86	1.80		
2. % of Latine students	5.15	3.00	0.35	1.72		
3. % of White students	-38.82	21.44	-1.15	-1.81		
4. % of FRL students	10.54	5.87	0.54	1.80		
5. % of IEP students	1.24	1.12	0.06	1.11		

Table 8 (cont'd)

1 & 2	-356.46	182.94	-1.33	2.87
1 & 3	153.89	1383.92	-	-
2 & 3	146.69	219.72	-	-

Enrollment Demographics^b

1. % of Black students	32.39***	1.32	0.87***	0.01
2. % of Latine students	1.57*	0.66	0.11*	0.04
3. % of White students	-33.85***	1.07	-	-
4. % of FRL students	12.07***	0.79	0.61***	0.03
5. % of IEP students	1.67	0.88	0.09	0.05
1 & 2	-241.56***	16.45	-0.87***	0.01
1 & 3	-	-	-	-
2 & 3	-	-	-	-

Mobility Rate

1. Mobility rate of all students	20.71***	0.69	0.98***	0.01
2. Mobility rate of Black students	20.20***	0.84	0.87***	0.01
3. Mobility rate of Latine students	22.39***	1.41	0.69***	0.03
4. Mobility rate of White students	19.77***	1.19	0.69***	0.03
5. Mobility rate of FRL students	22.18***	0.85	0.91***	0.01
6. Mobility rate of IEP students	20.02***	0.89	0.94***	0.01

School Staff per 100 students^a

1. Number of non-instructional paras	0.54***	0.08	0.32***	0.05
2. Number of instructional paras	2.60***	0.23	0.71***	0.05
3. Number of teachers	2.88***	0.21	1.03***	0.07

School Staff per 100 students^b

1. Number of non-instructional paras	0.55***	0.07	0.33***	0.04
2. Number of instructional paras	2.67***	0.15	0.72***	0.02
3. Number of teachers	2.80***	0.09	-	-

Support Staff per 100 students

1. Number of nurses	0.12***	0.01	0.48***	0.04
2. Number of psychologists	0.14***	0.01	0.90***	0.03
3. Number of social workers	0.32***	0.02	0.89***	0.03

Note. Indicators of good model fit: $\chi^2 p > .05$; CFI > .95; RMSEA < .05; SRMR < 0.08.

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

a. Model without negative residual variance constrained to zero.

b. Model with negative residual variance constrained to zero. Covariances with constructs with this set-to-zero residual variance are not estimated as a result.

Enrollment Demographics. The enrollment demographics construct includes the proportion of students that are Black, Latine, and White, and that qualify for FRL and have an IEP. The proportion of students in each racial/ethnic category were highly correlated with each other (i.e., as the enrollment of one group declines, the enrollment in other groups increases). Adding covariances among them improved CFA model fit. I included covariances among these variables as they are dependent upon each other (in this dataset, these categories are mutually exclusive, i.e., if a student is labeled as “White” they cannot be labeled as “Latine”). As the proportion of White students had a negative residual variance, I ran the model with and without this variance set to zero. Because standardized estimates are not available if a parameter has no residual variance and the remaining parameter estimates were similar, I used the standardized loadings from the model without the additional constraint to develop the weighted indicator, and report both in Table 8. The proportion of students that are Black, Latine, and qualify for FRL had positive and statistically significant factor loadings. The factor loading for the proportion of White students was negative and statistically significant. The factor loading for the proportion of students that qualify for an IEP was positive and non-significant. This means that in the subsequent models this indicator represents higher proportions of historically marginalized students enrolled.

Mobility. The mobility construct included the mobility rate of: all students, Black students, Latine students, White students, students that qualify for FRL, and students with an IEP. All mobility rates had positive and statistically significant factor loadings, with overall mobility rate having the strongest factor loading. Given that factor loadings did not differ by student subgroup—as anticipated—I did not create a weighted indicator for mobility and instead used the mobility rate of all students as an independent predictor in the subsequent models.

School Staff. The school staff factor included the number of non-instructional paraeducators, instructional paraeducators, and teachers per 100 students. Because there were only three variables, the model was just identified with zero degrees of freedom, and I could not estimate indices of model fit. As the number of teachers had a negative residual variance, I ran the model with and without this variance set to zero. Because standardized estimates are not available if a parameter has no residual variance and the remaining parameter estimates were similar, I used the standardized loadings from the model without the additional constraint to develop the weighted indicator, and report both in Table 8. All three variables had positive and statistically significant factor loadings. This means that in the subsequent models this indicator represents higher school staff-to-student ratios.

Support Staff. The support staff factor included the number of nurses, psychologists, and social workers per 100 students. With three variables, the model was just identified with zero degrees of freedom, and I could not estimate indices of model fit. All three variables had positive and statistically significant factor loadings. This means that in the subsequent models this indicator represents higher support staff-to-student ratios.

Multilevel Models

I estimated a series of latent growth curve models in Mplus Version 8.7 (Muthén & Muthén, 2021) using Maximum Likelihood with robust standard errors (MLR) to answer my research questions. MLR is an appropriate estimation method for this data as it is robust to missing and non-normal data (e.g., skewed or kurtotic distributions; Muthén & Muthén, 2021). In latent growth curve modeling, observations over time are used to estimate latent growth curve factors (i.e., the intercept and slope). I used a multilevel approach because the data is nested within three levels: (1) time – observations within schools; (2) school – observations across

schools; and (3) district – observations across districts. Using a multilevel approach, I was able to partition the variance accounted for by differences across schools (i.e., the within-level variance) and differences across districts (i.e., between-level variance) and examine how additional variables impact growth within and across levels. Unless otherwise noted, I followed a conventional multilevel modeling approach and constrained the residual variances of the outcome variables to be equal over time in the within part of the models, and in all models, I fixed the residual variances of the outcome variables to zero in the between part (Muthén & Muthén, 2021).

To answer my first research question (*How do partnership district schools—relative to their matched comparisons—experience changes in student outcomes over time?*), I first ran an unconditional “null” growth model of each student outcome. Then, I ran a multilevel growth model of each student outcome to account for schools being clustered within districts. Last, I ran a conditional multilevel growth model of each student outcome with partnership status predicting the intercept and slope at the between level, to determine whether there were differences in student outcomes over time by partnership status, while accounting for clustering. Below, I present simplified equations that represent this third model, using attendance rate as the example dependent variable.

Level 1: Time

$$Y_{ijt} = \beta_{0ij} + \beta_{1ij}\text{Time}_{tj} + e_{tij}$$

Where Y_{ijt} is school i in district j 's attendance rate at time t

Level 2: School

$$\beta_{0ij} = \delta_{00j} + u_{0ij}$$

Where β_{0ij} is the average attendance rate for school i in district j .

$$\beta_{1ij} = \delta_{10j} + u_{1ij}$$

Where β_{1ij} is the change in attendance rate for one school year for school i in district j .

Level 3: District

$$\delta_{00j} = \gamma_{000} + \gamma_{001}\text{Partnership}_j + v_{00j}$$

Where δ_{00j} is the average attendance rate for district j , when accounting for partnership status.

$$\delta_{10j} = \gamma_{101}\text{Partnership}_j + v_{10j}$$

Where δ_{10j} is the change in attendance rate for one school year for district j , when accounting for partnership status.

To answer my second research question, (*How do proxies for structural factors predict student outcomes?*), I ran an additional model for each student outcome, building upon the nested models I ran for the first research question. In these additional models, I added in the variables that represent proxies for structural factors as predictors of the intercept and slope at the within or between level, while accounting for SQI scores and partnership status. I grand mean centered all predictor variables except partnership status, which I effect coded (partnership status = 1; matched comparison status = -1).

Given my sample size, I modeled each predictor variable as non-time-varying (i.e., each predictor represents the average of all observed values over time) and used a staggered process to examine structural factors. First, I ran a model with all school-level variables at the within level. Second, I ran a model with all district-level predictors at the between level. Third, I ran a model with the statistically significant school-level predictors at the within level and the statistically significant district-level predictors at the between level. Finally, I entered all statistically

significant predictors from the third level into a final model.¹⁴ In all models I accounted for SQI scores and partnership status. Below, I present simplified equations that represent a model including enrollment and resources as the structural factors with attendance rate as the example dependent variable.

Level 1: Time

$$Y_{ijt} = \beta_{0ij} + \beta_{1ij}\text{Time}_{tj} + e_{tij}$$

Where Y_{ijt} is school i in district j 's attendance rate at time t .

Level 2: School

$$\beta_{0ij} = \delta_{00j} + \delta_{01j}\text{SQI} + \delta_{02j}\text{Resources} + \delta_{03j}\text{Enrollment} + u_{0ij}$$

Where β_{0ij} is the average attendance rate for school i in district j , when accounting for school-level SQI scores, resources, and enrollment.

$$\beta_{1ij} = \delta_{10j} + \delta_{11j}\text{SQI} + \delta_{12j}\text{Resources} + \delta_{13j}\text{Enrollment} + u_{1ij}$$

Where β_{1ij} is the change in attendance rate for one school year for school i in district j .

Level 3: District

$$\delta_{00j} = \gamma_{000} + \gamma_{001}\text{Partnership} + v_{00j}$$

Where δ_{00j} is the average enrollment for district j , when accounting for SQI, enrollment, resources, and partnership status.

$$\delta_{10j} = \gamma_{100} + \gamma_{102}\text{Partnership} + v_{10j}$$

Where δ_{10j} is the change in attendance rate for one school year for district j , when accounting for partnership status.

$$\delta_{20j} = \gamma_{200} + v_{20j}$$

¹⁴ I also ran a model with all variables for each outcome to see whether the direction, significance, or size of the parameter estimates from the large model differ from my approach. Unsurprisingly, these models tended to have poor model fit given that the number of parameters was larger than the number of clusters. As such, I only provide results for the models described above, but describe any key differences in my results writeup.

Where δ_{20j} is the change in attendance rate by SQI for district j .

$$\delta_{30j} = \gamma_{300} + v_{30j}$$

Where δ_{30j} is the change in attendance rate by resources for district j .

$$\delta_{40j} = \gamma_{400} + v_{40j}$$

Where δ_{40j} is the change in attendance rate by enrollment for district j .

Finally, to answer my third research question, (*Are the patterns identified in RQ1 and RQ2 the same when examining outcomes for different groups of students?*), I ran the same series of models outlined above for each student subgroup's outcomes. I only report findings for math growth percentiles, as it was the only outcome where all models successfully ran for each student subgroup. I encountered many convergence issues for the other outcomes due to small sample sizes and a lack of within cluster variation for some student subgroups.

MLM Fit Indices. I used the following four indices and respective critical values of adequate model fit to assess the fit of the multilevel models: (1) Chi-Square, $\chi^2 p > .05$ (Hu & Bentler, 1999; Mulaik et al., 1989); (2) Comparative Fit Index (CFI), $CFI > .90$ (Bentler, 1990; Hu & Bentler, 1999); (3) Tucker–Lewis index (TLI), $TLI > .90$ (Hu & Bentler, 1999); and (4) Root Mean Square Error of Approximation (RMSEA), $RMSEA < .08$ (Hu & Bentler, 1999; Steiger, 1998).

RESULTS

Descriptive Statistics

In the sections below, I provide descriptive statistics for each variable across time and districts, as well as across time broken down by partnership status. I provide supplemental tables that include descriptive statistics for each variable by school year and partnership status in Appendix C.

Measured Indicators of Structural Factors

I present descriptive statistics for the measured indicators of structural factors broken down by partnership status below in Table 9.

Enrollment. Across all years and schools, an average of 483 students ($SD = 352.91$) were enrolled in schools. Enrolled students were 46.63% Black ($SD = 63.41\%$), 36.12% White ($SD = 33.02$), and 11.56% Latine ($SD = 17.89\%$); 73.89% qualified for FRL ($SD = 17.83\%$) and 16.74% had an IEP ($SD = 16.39\%$). Partnership district schools had fewer overall students enrolled and lower proportions of White students enrolled, but higher proportions of Black students, Latine students, students with an IEP, and students who qualified for FRL enrolled compared to matched comparison district schools (see Table 9).

Mobility. Across all years and schools, the average school enrollment mobility rate was 16.18% for all students ($SD = 15.82\%$), 18.56% for Black students ($SD = 17.63\%$), 17.89% for White students ($SD = 21.64\%$), 15.17% for Latine students ($SD = 21.84\%$), 17.64% for students who qualify for FRL ($SD = 16.98\%$), and 25.13% for students with an IEP ($SD = 16.17\%$). Partnership district schools had higher mobility rates for all students and most student subgroups compared to matched comparison district schools; the mobility rate of students with an IEP was comparable between the two sets of schools (see Table 9).

School of choice enrollment was only available at the district level. Across all years and districts, districts lost an average of 0.53 resident students ($SD = 0.44$) and gained an average of 0.13 non-resident students ($SD = 0.13$) per student enrolled. Partnership districts had higher rates of resident students leaving compared to matched comparison districts; the rate of non-resident students arriving was comparable between the two sets of districts (see Table 9).

Academic Staff. Across all years and schools, schools had an average of 2.62 non-instructional paraeducators ($SD = 2.19$), 2.23 instructional paraeducators ($SD = 4.01$), and 6.91 teachers ($SD = 3.42$) per 100 students. Partnership district schools had higher numbers of instructional paraeducators per 100 students compared to matched comparison district schools; the ratios of non-instructional paraeducators and teachers to students were comparable between the two sets of schools (see Table 9).

Support Staff. Across all years and schools, schools had an average of 0.21 counselors ($SD = 2.39$) per 100 students. From 2015-16 to 2017-18, schools had an average of 0.04 nurses ($SD = 0.19$), 0.05 psychologists ($SD = 0.25$), and 0.15 social workers ($SD = 0.53$) per 100 students. Partnership district schools had higher numbers of counselors, but fewer social workers per 100 students compared to matched comparison district schools; the ratios of nurses and psychologists to students were comparable between the two sets of schools (see Table 8).

District Financial Characteristics. Across all districts from 2011-12 to 2018-19, districts earned an average of \$12,261.81 per student ($SD = \$4,500.57$), spent an average of \$12,089.34 per student ($SD = \4453.90), and had an average of -\$4.89 in their fund balance per student ($SD = \$2292.64$). Partnership districts had higher average revenue and expenditures per student, but a lower fund balance per student compared to matched comparison districts (see Table 9).

Table 9*Measured Indicators of Structural Factors by Partnership Status*

	Partnership District Schools		Matched Comparison District Schools	
	<u>Range</u>	<u>Mean (SD)</u>	<u>Range</u>	<u>Mean (SD)</u>
Total Enrollment	(31,2458)	462.31 (314.66)	(31,2790)	515.75 (404.56)
<u>Demographics</u>				
% Black	(0.00,100)	65.29 (32.39)	(0.00,79.83)	16.82 (17.95)
% Latine	(0.00,94.96)	13.18 (20.73)	(0.00,72.15)	8.97 (11.60)
% White	(0.00,88.71)	16.72 (20.92)	(4.24,100)	67.11 (23.88)
% FRL	(18.43,100)	79.95 (14.27)	(16.05,100)	64.21 (18.67)
% IEP	(0.47,100)	18.65 (17.49)	(2.28,100)	13.56 (13.80)
<u>Mobility Rate</u>				
All students	(0.00,100)	19.20 (16.19)	(0.00,100)	11.50 (14.00)
Black students	(0.00,100)	20.24 (17.03)	(0.00,100)	15.96 (18.24)
Latine students	(0.00,100)	22.69 (24.46)	(0.00,100)	11.07 (14.30)
White students	(0.00,100)	17.29 (24.27)	(0.00,100)	12.28 (17.60)
FRL students	(0.00,100)	20.40 (17.54)	(0.00,100)	13.40 (15.14)
IEP students	(4.85,100)	25.80 (15.23)	(5.26,100)	22.96 (18.75)
<u>School of Choice</u>				
% Residents leave	(0.14,2.47)	0.79 (0.48)	(0.07,0.80)	0.26 (0.16)
% Non-residents arrive	(0.00,0.63)	0.14 (0.16)	(0.01,0.36)	0.12 (0.10)
<u>Academic Staff</u>				
Non-instructional paras	(0.20,30.77)	2.79 (1.65)	(0.07,37.06)	2.34 (2.82)
Instructional paras	(0.04,32.55)	2.48 (3.90)	(0.00,86.82)	1.81 (4.14)
Teachers	(0.11,75.63)	7.00 (3.49)	(0.38,40.54)	6.77 (3.31)
<u>Support Staff</u>				
Counselors	(0.04,99.28)	0.25 (3.90)	(0.00,4.51)	0.14 (0.32)
Nurses	(0.00,1.94)	0.05 (0.18)	(0.00,2.97)	0.03 (0.20)
Psychologists	(0.00,0.50)	0.03 (0.08)	(0.00,6.93)	0.08 (0.37)
Social workers	(0.00,1.38)	0.10 (0.16)	(0.00,11.88)	0.22 (0.79)
<u>Financial Characteristics</u>				
Revenue ^a	(\$9,138.50, \$36,190.66)	\$14,124.27 (5,721.66)	(\$8185.07, \$12,591.04)	\$10,399.34 (\$987.67)
Expenditures	(\$8,910.35, \$36,393.80)	\$13,857.56 (5,710.19)	(\$8,278.36, \$12,489.04)	\$10,321.11 (\$978.72)

Table 9 (cont'd)

Fund Balance	(-\$10024.74, \$6195.20)	-\$542.84 (\$2786.03)	(-\$1,115.71, \$3726.73)	\$825.54 (\$503.62)
a. The district with the largest revenue per student (\$36,190.66) only served 511-552 students per year over time.				

Student Outcomes

I present descriptive statistics for student outcome variables broken down by partnership status below in Table 10.

Attendance. Across all schools from 2011-12 to 2018-19, the average attendance rate was 89.68% for the overall student population ($SD = 8.97\%$), 88.64% for Black students ($SD = 9.55\%$), 88.95% for White students ($SD = 14.78\%$), 87.34% for Latine students ($SD = 19.72\%$), 89.24% for FRL students who qualify for FRL ($SD = 8.93\%$), and 88.06% for students with an IEP ($SD = 10.17\%$). Partnership district schools had lower attendance rates for all students and student subgroups compared to matched comparison district schools (see Table 10).

Assessment Scores. To account for the difference in scoring across grade levels, I calculated z scores within grades and aggregated them to generate school and student subgroup-level averages. For the ELA portion of the M-STEP, across school years from 2014-15 to 2018-19, the average z score was 0.11 for the overall student population ($SD = 0.92$), -0.43 for Black students ($SD = 0.68$), 0.77 for White students ($SD = 0.88$), 0.12 for Latine students ($SD = 0.62$), -0.05 for students with an FRL ($SD = 0.82$), and -1.03 for students with an IEP ($SD = 0.62$). For the math portion of the M-STEP, across school years from 2014-15 to 2018-19, the average z score was 0.10 for the overall student population ($SD = 0.93$), -0.50 for Black students ($SD = 0.65$), 0.79 for White students ($SD = 0.84$), 0.16 for Latine students ($SD = 0.62$), -0.06 for students who qualify for FRL ($SD = 0.83$), and -1.02 for students with an IEP ($SD = 0.67$). Partnership district schools had lower and more negative ELA and math z scores (i.e., less than

the mean) for all students and student subgroups compared to matched comparison district schools (see Table 10).

Assessment Growth. For ELA, across school years from 2015-16 to 2018-19, the average growth percentile was 45.41 for the overall student population ($SD = 8.96$), 43.00 for Black students ($SD = 8.66$), 48.50 for White students ($SD = 9.11$), 47.50 for Latine students ($SD = 8.75$), 44.83 for students who qualify for FRL ($SD = 8.85$), and 41.72 for students with an IEP ($SD = 9.53$). For math, across school years from 2015-16 to 2018-19, the average growth percentile was 44.61 for the overall student population ($SD = 9.33$), 42.36 for Black students ($SD = 8.71$), 47.64 for White students ($SD = 9.96$), 46.60 for Latine students ($SD = 9.54$), 44.17 for students who qualify for FRL ($SD = 9.10$), and 42.28 for students with an IEP ($SD = 10.10$). Partnership district schools had lower ELA and math growth percentiles for all students and some student subgroups compared to matched comparison district schools; ELA growth percentiles for Black, White, and Latine students and math growth percentiles for Latine students were comparable between the two sets of schools (see Table 10).

Expulsions. Across all schools and years, an average of 0.15% of students were expelled ($SD = 0.66\%$). Partnership district schools had lower expulsions rates compared to matched comparison district schools (see Table 10).

Suspensions. Across all schools from 2009-10 to 2017-18, the average proportion of students who received out-of-school suspensions was 14.87% for the overall student population ($SD = 15.43\%$), 20.26% for Black students ($SD = 18.92\%$), 10.70% for White students ($SD = 15.17\%$), 9.68% for Latine students ($SD = 15.12\%$), and 19.96% for IEP students ($SD = 20.09\%$). Across all schools from 2009-10 to 2017-18, the average proportion of students who received in school suspensions was 2.52% for the overall student population ($SD = 7.19\%$),

3.81% for Black students ($SD = 10.25\%$), 2.11% for White students ($SD = 6.36\%$), 2.03% for Latine students ($SD = 6.62\%$), and 3.29% for IEP students ($SD = 9.10\%$). Partnership district schools had lower in-school suspensions rates for all students and all student subgroups and higher out-of-school suspensions rates for all students and all student subgroups compared to matched comparison district schools (see Table 10).

Table 10

Student Outcomes by Partnership Status

	Partnership District Schools		Matched Comparison District Schools	
	<u>Range</u>	<u>Mean (SD)</u>	<u>Range</u>	<u>Mean (SD)</u>
<u>Attendance Rate</u>				
All students	(6.77,100)	87.39 (8.50)	(0.00,100)	93.02 (8.58)
Black students	(0.00,100)	86.81 (8.77)	(0.00,100)	91.62 (10.00)
Latine students	(0.00,100)	83.93 (23.26)	(0.00,100)	91.99 (11.98)
White students	(0.00,100)	85.39 (17.61)	(0.00,100)	93.22 (8.62)
FRL students	(6.37,100)	87.05 (8.47)	(0.00,100)	92.46 (8.59)
IEP students	(0.00,100)	85.75 (9.84)	(0.00,100)	91.49 (9.66)
<u>ELA Z Scores</u>				
All students	(-2.54,3.30)	-0.33 (0.81)	(-1.68,3.40)	0.74 (0.68)
Black students	(-2.67,2.82)	-0.56 (0.67)	(-1.56,1.23)	-0.06 (0.54)
White students	(-1.49,3.36)	0.34 (1.00)	(-0.81,3.38)	1.02 (0.67)
Latine students	(-1.59,2.55)	-0.04 (0.65)	(-0.89,1.80)	0.38 (0.47)
FRL students	(-2.49,3.42)	-0.40 (0.75)	(-1.70,3.63)	0.45 (0.64)
IEP students	(-2.65,0.30)	-1.34 (0.47)	(-1.84,1.59)	-0.61 (0.54)
<u>Math Z Scores</u>				
All students	(-2.55,2.97)	-0.36 (0.78)	(-1.78,3.46)	0.76 (0.69)
Black students	(-2.62,2.59)	-0.63 (0.64)	(-2.13,1.06)	-0.14 (0.50)
FRL students	(-2.58,2.94)	-0.42 (0.72)	(-1.96,3.56)	0.49 (0.67)
IEP students	(-2.78,0.74)	-1.36 (0.51)	(-2.34,1.60)	-0.58 (0.60)
Latine students	(-1.58,1.96)	-0.01 (0.62)	(-0.61,2.07)	0.43 (0.50)
White students	(-1.60,3.36)	0.33 (0.94)	(-1.15,3.47)	1.06 (0.64)
<u>ELA Growth Percentiles</u>				
All students	(11.30,73.90)	43.58 (8.37)	(19.90,75.20)	48.15 (9.11)
Black students	(11.30,79.00)	42.29 (8.30)	(12.50,70.40)	44.71 (9.26)
White students	(20.80,82.70)	47.01 (9.39)	(11.90,75.90)	49.56 (8.77)
Latine students	(23.50,80.00)	47.60 (8.24)	(20.20,70.90)	47.38 (9.40)

Table 10 (cont'd)

FRL students	(7.10,73.60)	43.23 (8.35)	(21.50,72.50)	47.24 (9.04)
IEP students	(12.30,73.20)	40.16 (9.11)	(7.20,78.10)	44.18 (9.68)
<u>Math Growth Percentiles</u>				
All students	(14.50,73.50)	42.33 (8.51)	(19.60,77.40)	48.01 (9.49)
Black students	(15.60,77.70)	41.29 (8.25)	(18.50,72.20)	44.92 (9.25)
White students	(16.60,82.20)	45.15 (10.15)	(21.80,78.10)	49.43 (9.44)
Latine students	(23.80,76.70)	46.01 (9.04)	(17.50,79.20)	47.38 (10.15)
FRL students	(15.90,74.40)	42.09 (8.39)	(18.60,75.00)	47.30 (9.24)
IEP students	(10.10,80.60)	40.24 (9.41)	(17.30,83.00)	45.48 (10.33)
Expulsions	(0.00,13.21)	0.13 (0.55)	(0.00,18.92)	0.20 (0.80)
<u>In-School Suspensions^a</u>				
All students	(0.00,74.45)	1.67 (5.66)	(0.00,90.53)	9.24 (11.81)
IEP students	(0.00,82.35)	2.07 (6.66)	(0.00,88.24)	12.85 (15.51)
White students	(0.00,66.67)	1.42 (5.32)	(0.00,78.69)	7.73 (10.57)
Black students	(0.00,81.59)	2.17 (6.97)	(0.00,93.75)	15.86 (17.82)
Latine students	(0.00,55.56)	1.00 (4.28)	(0.00,80.00)	8.15 (14.13)
<u>Out-of-School Suspensions^a</u>				
All students	(0.00,98.06)	18.74 (16.42)	(0.00,95.42)	3.75 (8.81)
IEP students	(0.00,97.83)	24.86 (21.38)	(0.00,93.75)	5.06 (11.56)
White students	(0.00,87.59)	12.94 (17.55)	(0.00,91.49)	3.04 (7.45)
Black students	(0.00,94.89)	23.22 (19.07)	(0.00,96.30)	6.23 (13.36)
Latine students	(0.00,80.00)	10.90 (15.75)	(0.00,50.85)	3.35 (8.58)

- a. A number of schools had extremely large proportions of in- and out-of-school suspensions. Due to my liberal thresholds for implausible data, they were retained for analysis. There were no clear patterns to these extreme values, and they were distributed across both types of districts. As such, I describe them as a potential limitation to the data quality in my discussion.

School Characteristics

SQI Scores. From 2017-18 to 2018-19, the average SQI score was 49.79 ($SD = 23.47$). Partnership district schools had an overall average SQI of 39.76 ($SD = 19.03$) and matched comparison district schools had an overall average SQI of 64.26 ($SD = 21.63$). Partnership district schools had lower SQI scores than matched comparison district schools. In 2017-18, partnership district schools had an average SQI of 38.99 ($SD = 19.83$) and matched comparison

district schools had an average SQI of 62.40 ($SD = 23.56$). In 2018-19, partnership district schools had an average SQI of 40.20 ($SD = 20.39$) and matched comparison district schools had an average SQI of 62.24 ($SD = 24.90$).

RQ1: How do Partnership Districts—relative to their Matched Comparisons—experience changes in student outcomes over time?

In the sections below, I present the results of the multilevel growth models I ran to answer my first research question. Unless otherwise noted, I employed the recommended constraints that the residual variances of the outcome variables are equal over time in the within part of the model and zero in the between part of the model (Muthén & Muthén, 2021). For the multilevel models, I provide Intraclass Correlation Coefficients (ICCs), which—in the case of this study—indicate how similar schools within a district are on the outcome of interest.

Attendance

The residual variances for attendance at time points six, eight, and nine were too different in size compared to the remaining time points. Following established guidelines (Muthén & Muthén, 2021), I freely estimated the within-level residual variance at these three time points and constrained the remaining time points to be equivalent. I followed this approach in all attendance rate growth models.

For the RQ1 growth models of attendance rate, indices of model fit only met thresholds of adequate fit for RMSEA. Accounting for schools being nested within districts appeared to worsen model fit (Table 11) and the ICCs at each time point in the multilevel models were low (range = [.063,.202]). As such, I recommend interpreting the parameter estimates with caution.

Table 11*RQ1 Attendance Growth Models*

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	508	9	$X^2(35) = 105.66^{***}$	0.84	0.87	0.06
Multilevel	508	12	$X^2(68) = 247.15^{***}$	0.75	0.79	0.07
Partnership Status	508	14	$X^2(74) = 309.74^{***}$	0.71	0.75	0.08
Parameter Estimates						
<i>Within Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})		88.65***	-	-		
<u>Effects on Growth</u>						
Slope (b_{sw})		-0.19**	-	-		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)		47.52***	129.22**	129.29**		
Slope Variance (σ_{sw}^2)		0.75*	0.74*	0.74*		
Covariance (cov_{isw})		-4.97	-5.24	-5.22		
<i>Between Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{ib})		-	89.47***	89.52***		
Partnership Status (b_{ib1})		-	-	-1.31		
<u>Effects on Growth</u>						
Slope (b_{sb})		-	-0.20	-0.20		
Partnership Status (b_{sb1})		-	-	0.04		
<u>Random Effects</u>						
Intercept Variance (σ_{ib}^2)		-	13.52***	10.84		
Slope Variance (σ_{sb}^2)		-	0.01	0.01		
Covariance (cov_{isb})		-	0.01	0.08		

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

For all RQ1 attendance growth models, the within-level intercept and slope variance were statistically significant, indicating there were changes over time in attendance rate within schools and between schools. The within-level covariance between the intercept and slope was not statistically significant in any growth models, indicating a school's average attendance rate was

not associated with its estimated change over time. For the multilevel growth models, only the between-level intercept variance was statistically significant for the model without partnership status, indicating—in this model—there were changes in attendance between districts, but not within. The remaining between-level random effects were not statistically significant, indicating little variation in the outcome over time among districts.

For the unconditional growth model, the effect of time on attendance was negative and statistically significant. Across all schools and years, the average attendance rate was 88.65%, and attendance rates were estimated to decrease by 0.19% each year.

For the multilevel growth model, ICCs ranged from .065 to .202 across the years. Overall, the effect of time on attendance rate was not statistically significant. After accounting for clustering, across all schools and years, the average attendance rate was 89.47%.

For the multilevel growth model that included partnership status, ICCs ranged from .063 to .205 across the years. Overall, the effect of time on attendance rate was not statistically significant. After accounting for clustering and partnership status, across all districts and years, the average attendance rate was 89.52%. Partnership status did not statistically significantly predict average attendance rate or linear change in attendance rates over time.

Assessment

ELA Scores. For the RQ1 growth models of ELA scores, indices of model fit inconsistently met thresholds of adequate fit (Table 12). In my analyses, I multiplied ELA z scores by 10 to facilitate interpretation of the parameter estimates.

Table 12*RQ1 ELA Score Growth Models*

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	312	6	$\chi^2(14) = 48.91^{***}$	0.96	0.97	0.09
Multilevel	312	9	$\chi^2(26) = 149.57^{***}$	0.92	0.94	0.13
Partnership Status	312	11	$\chi^2(29) = 166.76^{***}$	0.93	0.94	0.13
Parameter Estimates						
<i>Within Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})		2.16***	-	-		
<u>Effects on Growth</u>						
Slope (b_{sw})		-0.55***	-	-		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)		89.40***	48.42***	47.76***		
Slope Variance (σ_{sw}^2)		0.98***	0.73***	0.73***		
Covariance (cov_{isw})		-3.50***	-2.11***	-2.08***		
<i>Between Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{ib})		-	3.65**	3.64***		
Partnership Status (b_{ib1})		-	-	-4.71***		
<u>Effects on Growth</u>						
Slope (b_{sb})		-	-0.71***	-0.71***		
Partnership Status (b_{sb1})		-	-	0.11		
<u>Random Effects</u>						
Intercept Variance (σ_{ib}^2)		-	26.42***	7.70**		
Slope Variance (σ_{sb}^2)		-	0.24***	0.24**		
Covariance (cov_{isb})		-	-0.49***	-0.09		

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

For all RQ1 ELA score growth models, the within-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating there were changes over time in ELA scores within schools and between schools, and a school's average ELA scores were negatively associated with estimated change over time (i.e., lower average

scores were associated with higher rates of change). For the multilevel growth models, the between-level intercept variance and slope variance were both statistically significant, indicating there were changes over time in ELA scores within and between districts. The covariance between the two was only statistically significant for the multilevel growth model that did not include partnership status, indicating—in this model—a district’s average ELA scores were negatively associated with estimated change over time.

For the unconditional growth model, the effect of time on ELA scores was negative and statistically significant. Across all included schools and years, the average transformed ELA score was 2.16 and scores were estimated to decrease by 0.55 each year.

For the multilevel growth model, ICCs ranged from .311 to .409 across the years. Overall, the effect of time on ELA scores was negative and statistically significant. After accounting for clustering, across all included schools and years, the average transformed ELA score was 3.65 and scores were estimated to decrease by 0.71 each year.

For the multilevel growth model that included partnership status, ICCs ranged from .349 to .450 across the years. Overall, the effect of time on ELA scores was negative and statistically significant. After accounting for clustering and partnership status, across all included districts and years the average transformed ELA score was 3.64 and scores were estimated to decrease by 0.71 each year. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 4.71 decrease in a district’s average ELA score. Partnership status did not statistically significantly predict linear change in ELA scores over time.

Math Scores. For the RQ1 growth models of math scores, indices of model fit inconsistently met thresholds of adequate fit (Table 13). In my analyses, I multiplied math z scores by 10 to facilitate interpretation of the parameter estimates.

Table 13

RQ1 Math Score Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	312	6	$\chi^2(14) = 122.79^{***}$	0.94	0.95	0.16
Multilevel	312	9	$\chi^2(26) = 209.59^{***}$	0.95	0.97	0.15
Partnership Status	312	11	$\chi^2(29) = 269.64^{***}$	0.95	0.96	0.16
Parameter Estimates						
<i>Within Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})		1.47**	-	-		
<u>Effects on Growth</u>						
Slope (b_{sw})		-0.24***	-	-		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)		83.24***	44.65***	44.06***		
Slope Variance (σ_{sw}^2)		0.76***	0.55**	0.55**		
Covariance (cov_{isw})		-1.40*	-0.75	-0.72		
<i>Between Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{ib})		-	2.89*	2.87***		
Partnership Status (b_{ib1})		-	-	-4.61***		
<u>Effects on Growth</u>						
Slope (b_{sb})		-	-0.39**	-0.39**		
Partnership Status (b_{sb1})		-	-	-0.08		
<u>Random Effects</u>						
Intercept Variance (σ_{ib}^2)		-	25.85***	7.38**		
Slope Variance (σ_{sb}^2)		-	0.27**	0.28*		
Covariance (cov_{isb})		-	0.12	-0.08		

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

For all RQ1 math score growth models, the within-level intercept variance and slope variance were statistically significant, indicating there were changes over time in math scores within schools and between schools. The covariance between the two was statistically significant only in the unconditional growth model, indicating—in this model—a school’s average math scores were negatively associated with estimated change over time (i.e., lower average scores associated with higher rates of change). For the multilevel growth models, the between-level intercept variance and slope variance were statistically significant, indicating there were changes in math scores within and between districts.

For the unconditional growth model, the effect of time on math scores was negative and statistically significant. Across all included schools and years, the average transformed math score was 1.47 and scores were estimated to decrease by 0.24 each year.

For the multilevel growth model, ICCs ranged from .338 to .416 across the years. Overall, the effect of time on math scores was negative and statistically significant. After accounting for clustering, across all included schools and years, the average transformed math score was 2.89 and scores were estimated to decrease by 0.39 each year.

For the multilevel growth model that included partnership status, ICCs ranged from .386 to .448 across the years. Overall, the effect of time on math scores was negative and statistically significant. After accounting for clustering and partnership status, across all included districts and years, the average transformed math score was 2.87 and scores were estimated to decrease by 0.39 each year. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 4.61 decrease in a district’s average math score. Partnership status did not statistically significantly predict linear change in math scores over time.

ELA Growth Percentiles. The residual variances for ELA growth at time point nine was too different in size compared to the remaining time points. Following established guidelines (Muthén & Muthén, 2021), I freely estimated the within-level residual variance at this time point and constrained the remaining time points to be equivalent. I followed this approach in all ELA growth percentile growth models.

For the RQ1 growth models of ELA growth percentiles, only the unconditional model met thresholds of adequate model fit, and accounting for schools being nested within districts appeared to worsen model fit (Table 14). As such, I recommend interpreting the parameter estimates with caution.

Table 14

RQ1 ELA Growth Percentile Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	378	7	$\chi^2(7) = 26.03^{***}$	0.95	0.96	0.09
Multilevel	378	10	$\chi^2(14) = 83.59^{***}$	0.78	0.81	0.12
Partnership Status	378	12	$\chi^2(16) = 84.15^{***}$	0.81	0.81	0.11
Parameter Estimates						
<i>Within Model</i>		<u>Single -</u>	<u>Multilevel -</u>	<u>Multilevel –</u>		
		<u>Unconditional</u>	<u>Unconditional</u>	<u>Partnership Status</u>		
	<u>Effects on Mean</u>					
	Intercept (b_{iw})	44.25***	-	-		
	<u>Effects on Growth</u>					
	Slope (b_{sw})	0.50**	-	-		
	<u>Random Effects</u>					
	Intercept Variance (σ_{iw}^2)	58.51***	40.01***	39.99***		
	Slope Variance (σ_{sw}^2)	4.15***	2.48	2.47		
	Covariance (cov_{isw})	-6.74**	-4.10*	-4.11**		
<i>Between Model</i>		<u>Single -</u>	<u>Multilevel -</u>	<u>Multilevel –</u>		
		<u>Unconditional</u>	<u>Unconditional</u>	<u>Partnership Status</u>		
	<u>Effects on Mean</u>					
	Intercept (b_{ib})	-	44.71***	44.75***		
Partnership Status (b_{ib1})	-	-	-2.15*			

Table 14 (cont'd)

<u>Effects on Growth</u>			
Slope (b_{sb})	-	0.04	0.02
Partnership Status (b_{sb1})	-	-	0.25
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	13.50	9.13*
Slope Variance (σ_{sb}^2)	-	1.30	1.27*
Covariance (cov_{isb})	-	-1.10	-0.60

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

For all RQ1 ELA growth percentile growth models, the within-level intercept variance and covariance between the slope and intercept were statistically significant, indicating there were changes over time in ELA growth percentiles within schools and a school's average ELA growth percentile was negatively associated with estimated change over time (i.e., lower average percentiles associated with higher rates of change). In the unconditional growth model, the within-level slope variance was also statistically significant, indicating—in this model—there were changes over time in ELA growth percentiles between schools. For the multilevel growth models, only the model that included partnership status had statistically significant random effects at the between level. The between-level intercept and slope variance were statistically significant, indicating—in this model—there were changes in ELA growth percentiles over time within and between districts.

For the unconditional growth model, the effect of time on ELA growth percentiles was positive and statistically significant. Across all included schools and years, the average ELA growth percentile was 44.25% and percentiles were estimated to increase by 0.50% each year.

For the multilevel growth model, ICCs ranged from .100 to .367 across the years. Overall, the effect of time on ELA growth percentiles was not statistically significant. After

accounting for clustering, across all included schools and years, the average ELA growth percentile was 44.71%.

For the multilevel growth model that included partnership status, ICCs ranged from .103 to .369 across the years. Overall, the effect of time on ELA growth percentiles was positive and not statistically significant. After accounting for clustering and partnership status, across all included schools and years, the average ELA growth percentile was 44.75%. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 2.15% decrease in a district's average ELA growth percentile. Partnership status did not statistically significantly predict linear change in ELA growth percentiles over time.

Math Growth Percentiles. For the RQ1 growth models of math growth percentiles, indices of model fit met most thresholds of adequate fit (Table 15).

Table 15

RQ1 Math Growth Percentile Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	379	6	$\chi^2(8) = 27.27^{***}$	0.95	0.96	0.08
Multilevel	379	9	$\chi^2(15) = 50.29^{***}$	0.92	0.94	0.08
Partnership Status	379	11	$\chi^2(17) = 49.38^{***}$	0.92	0.93	0.07
Parameter Estimates						
<i>Within Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})		43.19***	-	-		
<u>Effects on Growth</u>						
Slope (b_{sw})		0.74***	-	-		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)		63.52***	44.61***	44.64***		
Slope Variance (σ_{sw}^2)		3.38***	1.92*	1.92*		
Covariance (cov_{isw})		-5.87**	-2.74	-2.77		

Table 15 (cont'd)

<i>Between Model</i>	<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>
<u>Effects on Mean</u>			
Intercept (b_{ib})	-	44.19***	44.28***
Partnership Status (b_{ib1})	-	-	-2.89*
<u>Effects on Growth</u>			
Slope (b_{sb})	-	0.06	0.05
Partnership Status (b_{sb1})	-	-	0.18
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	15.10**	6.95*
Slope Variance (σ_{sb}^2)	-	1.28	1.24
Covariance (cov_{isb})	-	-2.05	-1.47

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

For all RQ1 math growth percentile growth models, the within-level intercept variance and slope variance were statistically significant, indicating there were changes over time in math growth percentiles within schools and between schools. The covariance between the two was statistically significant only in the unconditional growth model, indicating—in this model—a school's average math growth percentiles were negatively associated with estimated change over time (i.e., lower average growth percentiles associated with higher rates of change). For the multilevel growth models, the between-level intercept was statistically significant, indicating there were changes over time in math growth percentiles within districts.

For the unconditional growth model, the effect of time on math growth percentiles was positive and statistically significant. Across all included schools and years, the average math growth percentile was 43.19% and percentiles were estimated to increase by 0.74% each year.

For the multilevel growth model, ICCs ranged from .093 to .226 across the years. Overall, the effect of time on math growth percentiles was positive and not statistically significant. After accounting for clustering, across all included schools and years, the average math growth percentile was 44.19%.

For the multilevel growth model that included partnership status, ICCs ranged from .097 to .235 across the years. Overall, the effect of time on math growth percentiles was not statistically significant. After accounting for clustering and partnership status, across all included schools and years, the average math growth percentile was 44.28%. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 2.89% decrease in a district's average math growth percentile. Partnership status did not statistically significantly predict linear change in math growth percentiles over time.

Discipline

Expulsions. The residual variances for expulsions at time one, two, three, and five were too different in size compared to the remaining time points. Following established guidelines (Muthén & Muthén, 2021), I freely estimated the within-level residual variance at these four time points and constrained the remaining time points to be equivalent. I followed this approach in all expulsions growth models.

For the RQ1 growth models of expulsions, indices of model fit only met the threshold of adequate fit for RMSEA (Table 16).

Table 16

RQ1 Expulsions Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	571	10	$X^2(55) = 156.07^{***}$	0.74	0.79	0.06
Multilevel	571	13	$X^2(107) = 405.10^{***}$	0.74	0.78	0.07
Partnership Status	571	15	$X^2(115) = 424.50^{***}$	0.74	0.78	0.07
Parameter Estimates						
<i>Within Model</i> <u>Effects on Mean</u>	<u>Single -</u> <u>Unconditional</u>		<u>Multilevel -</u> <u>Unconditional</u>		<u>Multilevel –</u> <u>Partnership Status</u>	
Intercept (b_{iw})	0.19		-		-	

Table 16 (cont'd)

<u>Effects on Growth</u>			
Slope (b_{sw})	>0.01	-	-
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	0.43	0.37	0.37
Slope Variance (σ_{sw}^2)	>0.01	>0.01	>0.01
Covariance (cov_{isw})	-0.02	-0.02	-0.02
<i>Between Model</i>	<u>Single -</u>	<u>Multilevel -</u>	<u>Multilevel –</u>
	<u>Unconditional</u>	<u>Unconditional</u>	<u>Partnership Status</u>
<u>Effects on Mean</u>			
Intercept (b_{ib})	-	0.29***	0.29***
Partnership Status (b_{ib1})	-	-	-0.02
<u>Effects on Growth</u>			
Slope (b_{sb})	-	-0.01	-0.01
Partnership Status (b_{sb1})	-	-	>0.01
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	0.08*	0.08*
Slope Variance (σ_{sb}^2)	-	>0.01**	>0.01**
Covariance (cov_{isb})	-	-0.01*	-0.01*

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

For all RQ1 expulsions growth models, the within-level intercept variance, slope variance, and covariance between the two were not statistically significant, indicating little variation in expulsions among schools over time. For the multilevel growth models, the between-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating there were changes over time in the proportion of students who received expulsions within districts and between districts, and a district's average proportion of students who received expulsions was negatively associated with estimated change over time (i.e., lower proportions of students expelled associated with higher rates of change).

For the unconditional growth model, the effect of time on expulsions was not statistically significant. Across all included schools and years, the average proportion of students who received expulsions was 0.19%.

For the multilevel growth model, ICCs ranged from .099 to .539 across the years. Overall, the effect of time on expulsions was not statistically significant. After accounting for clustering, across all included schools and years, the average proportion of students who received expulsions was 0.29%.

For the multilevel growth model that included partnership status, ICCs ranged from .097 to .394 across the years. Overall, the effect of time on expulsions was not statistically significant. After accounting for clustering and partnership status, across all included schools and years, the proportion of students who received expulsions was 0.29%. Partnership status did not statistically significantly predict average expulsions or linear change in expulsions over time.

In-School Suspensions. For the RQ1 growth models of in-school suspensions, indices of model fit only met the threshold of adequate fit for RMSEA (Table 17).

Table 17

RQ1 In-School Suspensions Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	559	6	$\chi^2(14) = 32.72^{***}$	0.72	0.80	0.05
Multilevel	559	9	$\chi^2(26) = 50.38^{**}$	0.78	0.83	0.04
Partnership Status	559	11	$\chi^2(29) = 58.60^{***}$	0.78	0.81	0.04
Parameter Estimates						
<i>Within Model</i>		<u>Single -</u> <u>Unconditional</u>	<u>Multilevel -</u> <u>Unconditional</u>	<u>Multilevel –</u> <u>Partnership Status</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})		2.44***	-			-
<u>Effects on Growth</u>						
Slope (b_{sw})		>0.01	-			-

Table 17 (cont'd)

<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	39.13**	25.65	25.68
Slope Variance (σ_{sw}^2)	0.52	0.28	0.28
Covariance (cov_{isw})	-3.63*	-2.34	-2.35
<i>Between Model</i>	<u>Single -</u>	<u>Multilevel -</u>	<u>Multilevel –</u>
	<u>Unconditional</u>	<u>Unconditional</u>	<u>Partnership Status</u>
<u>Effects on Mean</u>			
Intercept (b_{ib})	-	5.54***	5.52***
Partnership Status (b_{ib1})	-	-	-0.72
<u>Effects on Growth</u>			
Slope (b_{sb})	-	-0.30*	-0.29*
Partnership Status (b_{sb1})	-	-	0.09
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	36.67***	35.31**
Slope Variance (σ_{sb}^2)	-	0.44**	0.42*
Covariance (cov_{isb})	-	-3.18**	-3.04**

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

Only the unconditional growth model had statistically significant within-level random effects. The within-level intercept variance and covariance between the intercept and slope were statistically significant, indicating—in this model—the proportion of students who received in-school suspensions varied over time within schools and a school's average proportion of in-school suspensions was negatively associated with linear change over time (i.e., lower proportion of students suspended associated with higher rates of change). For the multilevel expulsion growth models, the between-level intercept variance, slope variance, and covariance were statistically significant, indicating there were changes over time in the proportion of students who received in-school suspensions within districts and between districts, and a district's average proportion of in-school suspensions was negatively associated with estimated change over time.

For the unconditional growth model, the effect of time on in-school suspensions was not statistically significant. Across all included schools and years, the average proportion of students who received in-school suspensions was 2.44%.

For the multilevel growth model, ICCs ranged from .267 to .459 across the years. Overall, the effect of time on in-school suspensions was negative and statistically significant. After accounting for clustering, across all included schools and years, the average proportion of students who received in-school suspensions was 5.54%, and in-school suspensions were estimated to decrease by 0.30% each year.

For the multilevel growth model that included partnership status, ICCs ranged from .248 to .455 across the years. Overall, the effect of time on in-school suspensions was negative and statistically significant. After accounting for clustering and partnership status, across all included schools and years, the average proportion of students who received in-school suspensions was 5.52%, and in-school suspensions were estimated to decrease by 0.29% each year. Partnership status did not statistically significantly predict average in-school suspensions or linear change in in-school suspensions over time.

Out-of-School Suspensions. The residual variances for out-of-school suspensions at time zero was too different in size compared to the remaining time points. Following established guidelines (Muthén & Muthén, 2021), I freely estimated the within-level residual variance at this time point and constrained the remaining time points to be equivalent. I followed this approach in all out-of-school suspensions growth models.

For RQ1 out-of-school suspension growth models, only the model including partnership status met the threshold of adequate fit for RMSEA (Table 18). As such, I recommend interpreting parameter estimates with caution.

Table 18*RQ1 Out-of-School Suspensions Growth Models*

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Single	559	7	$\chi^2(13) = 80.05^{***}$	0.83	0.87	0.10
Multilevel	559	10	$\chi^2(25) = 131.75^{***}$	0.72	0.78	0.09
Partnership Status	559	12	$\chi^2(28) = 132.52^{***}$	0.77	0.79	0.08
Parameter Estimates						
<i>Within Model</i>	<u>Single -</u>	<u>Multilevel -</u>	<u>Multilevel -</u>			
<u>Effects on Mean</u>	<u>Unconditional</u>	<u>Unconditional</u>	<u>Partnership Status</u>			
Intercept (b_{iw})	15.10***	-	-			
<u>Effects on Growth</u>						
Slope (b_{sw})	0.28*	-	-			
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	229.47***	200.58***	202.47***			
Slope Variance (σ_{sw}^2)	3.00***	2.32*	2.35*			
Covariance (cov_{isw})	-12.69***	-12.22**	-12.47***			
<i>Between Model</i>	<u>Single -</u>	<u>Multilevel -</u>	<u>Multilevel -</u>			
<u>Effects on Mean</u>	<u>Unconditional</u>	<u>Unconditional</u>	<u>Partnership Status</u>			
Intercept (b_{ib})	-	14.80***	13.98***			
Partnership Status (b_{ib1})	-	-	4.68***			
<u>Effects on Growth</u>						
Slope (b_{sb})	-	0.23	0.30			
Partnership Status (b_{sb1})	-	-	0.15			
<u>Random Effects</u>						
Intercept Variance (σ_{ib}^2)	-	32.98	4.87			
Slope Variance (σ_{sb}^2)	-	0.60	0.52			
Covariance (cov_{isb})	-	0.14	0.21			

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

For all RQ1 out-of-school suspension growth models, the within-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating there were changes over time in the proportion of students who received out-of-school suspensions within schools and between schools, and a school's average proportion of out-of-school

suspensions was negatively associated with estimated change over time (i.e., lower proportion of suspended students associated with higher rates of change). For the multilevel growth models, the between-level intercept variance, slope variance, and covariance were statistically significant, indicating there were changes over time in the proportion of students who received out-of-school suspensions within districts and between districts, and a district's average proportion of out-of-school suspensions was negatively associated with estimated change over time.

For the unconditional growth model, the effect of time on out-of-school suspensions was positive and statistically significant. Across all included schools and years, the average proportion of students who received out-of-school suspensions was 15.10%, and out-of-school suspensions were estimated to increase by 0.28% each year.

For the multilevel growth model, ICCs ranged from .146 to .308 across the years. Overall, the effect of time on out-of-school suspensions was not statistically significant. After accounting for clustering, across all included schools and years, the average proportion of students who received out-of-school suspensions was 14.80%.

For the multilevel growth model that included partnership status, ICCs ranged from .160 to .323 across the years. Overall, the effect of time on out-of-school suspensions was not statistically significant. After accounting for clustering and partnership status, across all included schools and years, the average proportion of students who received out-of-school suspensions was 13.98%. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 4.68% increase in a district's average proportion of out-of-school suspensions. Partnership status did not statistically significantly predict linear change in out-of-school suspensions over time.

Summary of RQ1 Findings

Partnership status statistically significantly predicted average ELA scores (negatively), math scores (negatively), ELA growth percentiles (negatively), math growth percentiles (negatively), and out-of-school suspensions (positively). Even after accounting for clustering and partnership status, ELA scores, math scores, and in-school suspension rates were statistically significantly declining over time across the sample. However, partnership status did not statistically significantly predict linear changes over time for any of the outcomes. Taken together, these findings suggest that partnership district schools had worse average student outcomes, but their trajectories over time did not differ from matched comparison district schools above what would be expected by chance.

RQ2: How do proxies for structural factors predict student outcomes?

In the sections below, I present the results of the multilevel growth models I ran to answer my second research question. To ensure I was consistent across models, I followed the same modeling conventions for each outcome variable (i.e., constraining residual variances to be equivalent or allowing some to vary) as I employed in RQ1.

Attendance

For the RQ2 growth models of attendance rate, indices of model fit only met thresholds of adequate fit for RMSEA (Table 19).

Table 19

RQ2 Attendance Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	331	28	$\chi^2(116) = 351.00^{***}$	0.78	0.77	0.08
District-level factors	400	26	$\chi^2(110) = 359.95^{***}$	0.78	0.77	0.08
Final Model	337	26	$\chi^2(110) = 325.42^{***}$	0.79	0.78	0.08

Table 19 (cont'd)

Parameter Estimates			
<i>Within Model</i>	<i>School Factors</i>	<i>District Factors</i>	<i>Final Model</i>
<u>Effects on Mean</u>			
SQI (b_{iw1})	0.02	0.21***	0.02
Enrollment (b_{iw2})	>0.01**	-	>0.01**
Demographics (b_{iw3})	-0.01	-	-
Mobility Rate (b_{iw4})	-0.39**	-	-0.40**
School Staff (b_{iw5})	-0.25	-	-0.17
Support Staff (b_{iw6})	0.81	-	-
Counselors (b_{iw7})	-0.06	-	-0.04
<u>Effects on Growth</u>			
SQI (b_{sw1})	0.01***	>0.01	>0.01*
Enrollment (b_{sw2})	>0.01	-	>0.01
Demographics (b_{sw3})	>0.01	-	-
Mobility Rate (b_{sw4})	>0.01	-	0.01
School Staff (b_{sw5})	0.01*	-	0.01
Support Staff (b_{sw6})	-0.04	-	-
Counselors (b_{sw7})	-0.01*	-	-0.02*
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	18.84***	61.08*	18.57***
Slope Variance (σ_{sw}^2)	0.31**	0.50**	0.30**
Covariance (cov _{isw})	-1.54*	-3.31	-1.46*
<i>Between Model</i>	<i>School Factors</i>	<i>District Factors</i>	<i>Final Model</i>
<u>Effects on Mean</u>			
Intercept (b_{ib})	91.73***	90.87***	92.39***
Partnership Status (b_{ib1})	-0.25	0.69	-0.77
Residents Out (b_{ib2})	-	-1.90	-
Non-Residents In (b_{ib3})	-	1.35	5.44
Expenditures (b_{ib4})	-	-0.50	-
Revenue (b_{ib5})	-	0.49	-
Fund Balance (b_{ib6})	-	0.02	-
<u>Effects on Growth</u>			
Slope (b_{sb})	-0.27*	-0.29***	-0.36***
Partnership Status (b_{sb1})	0.01	0.06	0.05
Residents Out (b_{sb2})	-	-0.19	-
Non-Residents In (b_{sb3})	-	-1.45*	-1.82**
Expenditures (b_{sb4})	-	0.05	-
Revenue (b_{sb5})	-	-0.04	-
Fund Balance (b_{sb6})	-	>0.01	-

Table 19 (cont'd)

<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	1.89	0.24	2.21
Slope Variance (σ_{sb}^2)	0.01	>0.01	>0.01
Covariance (cov _{isb})	0.05	0.03	0.01**

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

For all RQ2 attendance growth models, the within-level intercept variance and slope variance were statistically significant, indicating there were changes over time in attendance rates within schools and between schools. The within-level covariance between the slope and intercept was also statistically significant in the school-level factors model, indicating—in this model—a school's average attendance rate was negatively associated with estimated change over time (i.e., lower average attendance associated with higher rates of change). None of the between-level random effects were statistically significant except the covariance between the intercept and slope for the final model, indicating—in this model—districts' average attendance rates were negatively associated with estimated change over time. In the other two models, there was little variation in attendance rate among districts over time.

For the school-level factors model, the effect of time was negative and statistically significant. Across all included schools and years, with SQI scores and school-level factors at their average value, the average attendance rate was 91.73%, and attendance rates were estimated to decrease by 0.27% each year. Partnership status did not statistically significantly predict average attendance rates or their linear change over time. Statistically significant predictors of average attendance rates included enrollment change (positive) and mobility rate (negative). Statistically significant predictors of attendance rate change over time included SQI scores (positive), school staff-to-student ratios (positive), and counselor-to-student ratios (negative).

For the district-level factors model, the effect of time was negative and statistically significant. Across all included schools and years, with SQI scores and district-level factors at their average value, the average attendance rate was 90.87%, and attendance rates were estimated to decrease by 0.29% each year. Partnership status did not statistically significantly predict average attendance rates or their linear change over time. Only one district-level factor emerged as statistically significant; the proportion of non-resident students arriving negatively predicted attendance rate change over time.

For the final model, the effect of time was negative and statistically significant. Across all included schools and years, with SQI scores and included school- and district-level factors at their average value, the average attendance rate was 92.39%, and attendance rates were estimated to decrease by 0.36% each year. Partnership status did not statistically significantly predict average attendance rates or their linear change over time. Patterns of the predictors were the same as the previous models, except school staff-to-student ratios longer statistically significantly predicted attendance rate change over time.

Assessment

ELA Scores. For the RQ2 ELA score growth models, indices of model fit for the school-level factors model did not meet any thresholds of adequate fit. The indices of model fit for the district-level factors and final models inconsistently met model fit thresholds (Table 20). In my analyses, I multiplied ELA z scores by 10 to facilitate interpretation of the parameter estimates.

Table 20

RQ2 ELA Score Growth Models

<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Model Fit</u>			
			<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	263	25	$\chi^2(50) = 260.38^{***}$	0.89	0.87	0.13
District-level factors	303	23	$\chi^2(47) = 316.92^{***}$	0.92	0.90	0.14

Table 20 (cont'd)

Final Model	290	19	$\chi^2(41) = 240.27^{***}$	0.90	0.89	0.13
Parameter Estimates						
<i>Within Model</i>	<u>School Factors</u>		<u>District Factors</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
SQI (b_{iw1})	0.03***		0.30***		0.18***	
Enrollment (b_{iw2})	>0.01		-		-	
Demographics (b_{iw3})	-0.05**		-		-0.05*	
Mobility Rate (b_{iw4})	-0.15		-		-0.18*	
School Staff (b_{iw5})	-0.13		-		-	
Support Staff (b_{iw6})	0.30		-		-	
Counselors (b_{iw7})	8.86***		-		6.90**	
<u>Effects on Growth</u>						
SQI (b_{sw1})	0.03***		0.01*		0.03***	
Enrollment (b_{sw2})	>0.01		-		-	
Demographics (b_{sw3})	0.01		-		0.01	
Mobility Rate (b_{sw4})	0.04*		-		0.03*	
School Staff (b_{sw5})	-0.06		-		-	
Support Staff (b_{sw6})	0.33		-		-	
Counselors (b_{sw7})	0.51		-		0.60	
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	13.69***		22.36***		14.16***	
Slope Variance (σ_{sw}^2)	0.59***		0.69***		0.59***	
Covariance (cov _{isw})	-2.31***		-2.91***		-2.29***	
<i>Between Model</i>	<u>School-Level</u>		<u>District-Level</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
Intercept (b_{ib})	1.87**		3.33***		1.73**	
Partnership Status (b_{ib1})	-0.04		-1.07		0.06	
Residents Out (b_{ib2})	-		1.43		-	
Non-Residents In (b_{ib3})	-		3.24		-	
Expenditures (b_{ib4})	-		-0.34		-	
Revenue (b_{ib5})	-		0.32		-	
Fund Balance (b_{ib6})	-		0.08		-	
<u>Effects on Growth</u>						
Slope (b_{sb})	-0.65***		-0.74***		-0.66***	
Partnership Status (b_{sb1})	0.27		0.30		0.18	
Residents Out (b_{sb2})	-		-0.28		-	
Non-Residents In (b_{sb3})	-		-0.73		-	
Expenditures (b_{sb4})	-		0.03		-	
Revenue (b_{sb5})	-		-0.03		-	
Fund Balance (b_{sb6})	-		-0.01		-	

Table 20 (cont'd)

<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	6.52*	1.59	6.57*
Slope Variance (σ_{sb}^2)	0.20*	0.17	0.19*
Covariance (cov_{isb})	-0.88*	-0.50	-0.80*

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

For all RQ2 ELA score models, the within-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating there were changes over time in ELA scores within schools and between schools, and a school's average ELA scores were negatively associated with estimated change over time (i.e., lower average scores associated with higher rates of change). For the school-level factors and final models, the between-level intercept variance, slope variance, and covariance between the two were also statistically significant, indicating—in these models—there were changes over time in ELA scores within and between districts, and a district's average ELA scores were negatively associated with estimated change over time.

For the school-level factors model, the effect of time was negative and statistically significant. Across all included districts and years with SQI scores and school-level factors at their average value, the average transformed ELA score was 1.87 and scores were estimated to decrease by 0.65 each year. After including SQI scores and school-level factors, partnership status no longer statistically significantly predicted average ELA scores. Statistically significant school-level predictors of average ELA scores included SQI scores (positive), the latent factor of demographics representing the proportion of historically marginalized students (negative), and counselor-to-student ratios (positive). Statistically significant school-level predictors of ELA score change over time included SQI scores (positive) and mobility rate (positive).

For the district-level factors model, the effect of time was negative and statistically significant. Across all included districts and years with SQI scores and district-level factors at their average value, the average transformed ELA score was 3.33 and scores were estimated to decrease by 0.74 each year. Partnership status and district-level factors did not statistically significantly predict average ELA scores or their linear change over time.

Because none of the district-level factors were statistically significant, the final model included only school-level factors. The effect of time was negative and statistically significant. Across all included districts and years, with SQI scores and included school-level factors at their average value, the average transformed ELA score was 1.73 and scores were estimated to decrease by 0.66 each year. Partnership status did not statistically significantly predict average ELA scores or their linear change over time. Patterns of the predictors were the same as the school-level factors only model, except mobility rate emerged as a statistically significant negative predictor of average ELA scores.

Math Scores. For the RQ2 math score growth models, indices of model fit inconsistently met = thresholds of adequate fit (Table 21). In my analyses, I multiplied math z scores by 10 to facilitate interpretation of the parameter estimates.

Table 21

RQ2 Math Score Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	263	25	$\chi^2(50) = 320.64^{***}$	0.90	0.88	0.14
District-level factors	303	23	$\chi^2(47) = 470.85^{***}$	0.92	0.91	0.17
Final Model	290	21	$\chi^2(44) = 284.53^{***}$	0.91	0.90	0.14
Parameter Estimates						
<u>Within Model</u>	<u>School Factors</u>		<u>District Factors</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
SQI (b_{iw1})	0.18***		0.30***		0.18***	
Enrollment (b_{iw2})	>0.01		-		-	

Table 21 (cont'd)

Demographics (b_{iw3})	-0.05**	-	-0.05**
Mobility Rate (b_{iw4})	-0.18***	-	-0.20***
School Staff (b_{iw5})	-0.20	-	-
Support Staff (b_{iw6})	0.30	-	-
Counselors (b_{iw7})	6.75**	-	4.87*
<u>Effects on Growth</u>			
SQI (b_{sw1})	0.03***	0.02***	0.03***
Enrollment (b_{sw2})	>0.01	-	-
Demographics (b_{sw3})	>0.01*	-	0.01*
Mobility Rate (b_{sw4})	0.03	-	0.02
School Staff (b_{sw5})	-0.04	-	-
Support Staff (b_{sw6})	-0.57	-	-
Counselors (b_{sw7})	0.83	-	0.94
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	10.59***	18.06***	10.96***
Slope Variance (σ_{sw}^2)	0.38**	0.50**	0.42**
Covariance (cov _{isw})	-1.39***	-1.95**	-1.41***
<u>Between Model</u>			
<u>Effects on Mean</u>	<u>School-Level</u>	<u>District-Level</u>	<u>Final Model</u>
Intercept (b_{ib})	1.26***	2.34***	1.08*
Partnership Status (b_{ib1})	-0.14	-1.21**	-0.12
Residents Out (b_{ib2})	-	1.93	-
Non-Residents In (b_{ib3})	-	-1.87	-
Expenditures (b_{ib4})	-	-0.29	-
Revenue (b_{ib5})	-	0.28	-
Fund Balance (b_{ib6})	-	0.10*	0.01
<u>Effects on Growth</u>			
Slope (b_{sb})	-0.31*	-0.43***	-0.35**
Partnership Status (b_{sb1})	0.06	0.23	0.04
Residents Out (b_{sb2})	-	-0.36	-
Non-Residents In (b_{sb3})	-	-1.02	-
Expenditures (b_{sb4})	-	-0.03	-
Revenue (b_{sb5})	-	0.03	-
Fund Balance (b_{sb6})	-	-0.01	-0.01
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	2.61**	0.69	2.00
Slope Variance (σ_{sb}^2)	-0.31*	0.16	0.16
Covariance (cov _{isb})	-0.41	-0.33	-0.29

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

For all RQ2 math score growth models, the within-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating there were changes over time in math scores within schools and between schools, and a school's average math scores were negatively associated with estimated change over time (i.e., lower average scores associated with higher rates of change). Only the school-level factors model had statistically significant between-level random effects, indicating—in this model—there were changes over time in math scores within and between districts, and a district's average math scores were negatively associated with estimated change over time. In the other two models, there was little variation in math scores among districts over time.

For the school-level factors model, the effect of time was negative and statistically significant. Across all included districts and years, with SQI scores and school-level factors at their average value, the average transformed math score was 1.26 and scores were estimated to decrease by 0.31 each year. After including SQI scores and school-level factors, partnership status no longer statistically significantly predicted average math scores. Statistically significant school-level predictors of average math scores included SQI scores (positive), the latent factor of demographics representing the proportion of historically marginalized students (negative), mobility rate (negative), and counselor-to-student ratios (positive). Statistically significant school-level predictors of math score change over time included SQI scores (positive) and the latent factor of demographics representing the proportion of historically marginalized students (positive).

For the district-level factors model, the effect of time was negative and statistically significant. Across all included districts and years, with SQI scores and district-level factors at their average value, the average transformed math score was 2.34 and scores were estimated to

decrease by 0.43 each year. Statistically significant district-level predictors of average math scores included partnership status (negative) and district fund balance (positive).

For the final model, the effect of time was negative and statistically significant. Across all included schools and years, with SQI scores and included school- and district-level factors at their average value, the average transformed math score was 1.08 and scores were estimated to decrease by 0.35 each year. Partnership status did not statistically significantly predict average math scores or their linear change over time. District-level factors were no longer statistically significant predictors in the final model; patterns of the predictors were the same as the school-level factors only model.

ELA Growth Percentiles. For the RQ2 growth models of ELA growth percentiles, none of the indices of model fit met thresholds of adequate fit (Table 22). As such, I recommend interpreting the parameter estimates with caution.

Table 22

RQ2 ELA Growth Percentiles Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	315	26	$\chi^2(30) = 126.99^{***}$	0.84	0.76	0.10
District-level factors	371	24	$\chi^2(28) = 134.48^{***}$	0.87	0.81	0.10
Final Model	321	20	$\chi^2(24) = 111.09^{***}$	0.86	0.82	0.11
Parameter Estimates						
<u>Within Model</u>	<u>School Factors</u>		<u>District Factors</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
SQI (b_{iw1})	0.18***		0.23***		0.20***	
Enrollment (b_{iw2})	>0.01**		-		>0.01**	
Demographics (b_{iw3})	>0.01		-		-	
Mobility Rate (b_{iw4})	-0.16		-		-	
School Staff (b_{iw5})	0.17		-		-	
Support Staff (b_{iw6})	-1.32*		-		-0.80***	
Counselors (b_{iw7})	0.25		-		0.21	
<u>Effects on Growth</u>						
SQI (b_{sw1})	0.03		0.01		0.01	

Table 22 (cont'd)

Enrollment (b_{sw2})	>0.01	-	>0.01
Demographics (b_{sw3})	>0.01	-	-
Mobility Rate (b_{sw4})	0.05	-	-
School Staff (b_{sw5})	0.04	-	-
Support Staff (b_{sw6})	-0.07	-	0.11
Counselors (b_{sw7})	-0.09**	-	-0.08*
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	19.63***	22.22***	20.47***
Slope Variance (σ_{sw}^2)	2.54*	2.62	2.75*
Covariance (cov _{isw})	-4.73**	-4.43**	-4.72**
<u>Between Model Effects on Mean</u>			
	<u>School-Level</u>	<u>District-Level</u>	<u>Final Model</u>
Intercept (b_{ib})	44.82***	44.51***	44.79***
Partnership Status (b_{ib1})	-0.29	-0.23	-0.18
Residents Out (b_{ib2})	-	1.34	-
Non-Residents In (b_{ib3})	-	0.78	-
Expenditures (b_{ib4})	-	-0.22	-
Revenue (b_{ib5})	-	0.21	-
Fund Balance (b_{ib6})	-	>0.01	-
<u>Effects on Growth</u>			
Slope (b_{sb})	0.01	-0.04	0.02
Partnership Status (b_{sb1})	0.40	0.71	0.43
Residents Out (b_{sb2})	-	-0.87	-
Non-Residents In (b_{sb3})	-	-2.42	-
Expenditures (b_{sb4})	-	0.05	-
Revenue (b_{sb5})	-	-0.06	-
Fund Balance (b_{sb6})	-	-0.01	-
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	6.18**	3.81*	4.57*
Slope Variance (σ_{sb}^2)	1.43*	1.06	1.23*
Covariance (cov _{isb})	-1.60	-1.00	-0.95

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

For all RQ2 ELA growth percentile growth models, the between-level intercept, within-level intercept variance, and within-level covariance between the slope and intercept were statistically significant, indicating there were changes over time in ELA growth percentiles within schools, within districts, and a school's average ELA growth percentile was negatively

associated with estimated change over time (i.e., lower average percentiles associated with higher rates of change). For the school-level factors and final models, the within- and between-level slope variances were also statistically significant, indicating—in these models—there were changes over time in ELA growth percentiles between schools and between districts.

For the school-level factors model, time was not statistically significant. Across all included schools and years, with SQI scores and school-level factors at their average value, the average ELA growth percentile was 44.82. Time and partnership status were not statistically significant predictors of ELA growth percentiles. Statistically significant school-level predictors of average ELA growth percentiles included SQI scores (positive), enrollment change (positive), and school staff-to-student ratios (negative). Statistically significant school-level predictors of ELA growth percentiles change over time included counselor-to-student ratios (negative).

For the district-level factors model, across all included schools and years, with SQI scores and district-level factors at their average value, the average ELA growth percentile was 44.51. Time, partnership status, and district-level factors were not statistically significant predictors of ELA growth percentiles or their linear change over time.

Because none of the district-level factors were statistically significant, the final model includes only the statistically significant school-level factors. For the final model, across all included schools and years, with SQI scores and included school-level factors at their average value, the average ELA growth percentile was 44.79. Time and partnership status were not statistically significant predictors of ELA growth percentiles. Patterns of the predictors were the same as the school-level factors only model.

Math Growth Percentiles. For the RQ2 math growth percentiles growth models, the district-level factors model only met thresholds of adequate fit for CFI. Indices of model fit for

the school-level factors and final models met most model fit thresholds of adequate fit (Table 23).

Table 23

RQ2 Math Growth Percentiles Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	316	25	$\chi^2(31) = 91.85^{***}$	0.93	0.90	0.08
District-level factors	372	23	$\chi^2(29) = 122.49^{***}$	0.92	0.88	0.09
Final Model	368	23	$\chi^2(29) = 84.06^{***}$	0.94	0.92	0.07
Parameter Estimates						
<i>Within Model</i> <u>Effects on Mean</u>	<u>School Factors</u>		<u>District Factors</u>		<u>Final Model</u>	
SQI (b_{iw1})	0.22***		0.24***		0.20***	
Enrollment (b_{iw2})	>0.01		-		-	
Demographics (b_{iw3})	0.01		-		-	
Mobility Rate (b_{iw4})	-0.12*		-		-0.15*	
School Staff (b_{iw5})	0.21		-		-	
Support Staff (b_{iw6})	-1.41*		-		-0.49**	
Counselors (b_{iw7})	0.37*		-		0.33	
<u>Effects on Growth</u>						
SQI (b_{sw1})	0.02		0.01		0.01	
Enrollment (b_{sw2})	>0.01		-		-	
Demographics (b_{sw3})	>0.01		-		-	
Mobility Rate (b_{sw4})	0.03		-		0.03	
School Staff (b_{sw5})	0.04		-		-	
Support Staff (b_{sw6})	0.19		-		0.30***	
Counselors (b_{sw7})	-0.10***		-		-0.10***	
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	22.55***		24.16***		23.00***	
Slope Variance (σ_{sw}^2)	2.05*		1.84*		1.84*	
Covariance (cov_{isw})	-3.56*		-2.95*		-2.74	
<i>Between Model</i> <u>Effects on Mean</u>	<u>School-Level</u>		<u>District-Level</u>		<u>Final Model</u>	
Intercept (b_{ib})	44.35***		44.04***		43.92***	
Partnership Status (b_{ib1})	-0.65		-0.68		-0.71	
Residents Out (b_{ib2})	-		-1.99		-	
Non-Residents In (b_{ib3})	-		13.97*		8.57	
Expenditures (b_{ib4})	-		0.11		-	
Revenue (b_{ib5})	-		-0.06		-	

Table 23 (cont'd)

Fund Balance (b_{ib6})	-	-0.09*	-0.05
<u>Effects on Growth</u>			
Slope (b_{sb})	0.15	-0.11	>0.01
Partnership Status (b_{sb1})	0.14	0.31	0.25
Residents Out (b_{sb2})	-	1.26	-
Non-Residents In (b_{sb3})	-	-5.51	-4.05
Expenditures (b_{sb4})	-	0.06	-
Revenue (b_{sb5})	-	-0.08	-
Fund Balance (b_{sb6})	-	>0.01	-0.02
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	5.77*	2.95*	4.27*
Slope Variance (σ_{sb}^2)	1.24	0.65*	0.89
Covariance (cov_{isb})	-1.88	-0.96	-1.50

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

For all RQ2 math growth percentile models, the within-level intercept variance and slope variance and the between-level intercept variance were statistically significant, indicating there were changes over time in math growth percentiles within schools and districts, and between schools. For the school-level and district-level factors models, the within-level intercept and slope covariance were also statistically significant, indicating—in these models—a school's average math growth percentile was negatively associated with estimated change over time (i.e., lower average percentiles associated with higher rates of change). For the district-level factors model, the between-level slope variance was also statistically significant, indicating—in this model—there were changes over time in math growth percentiles between districts.

For the school-level factors model, across all included schools and years with SQI scores and school-level factors at their average value, the average math growth percentile was 44.35. Time and partnership status were not statistically significant predictors of math growth percentiles. Statistically significant school-level predictors of average math growth percentiles included SQI scores (positive), mobility rate (negative), support staff-to-student ratios (negative),

and counselor-to-student ratios (positive). Statistically significant school-level predictors of math growth percentiles change over time included counselor-to-student ratios (negative).

For the district-level factors model, across all included schools and years with SQI scores and district-level factors at their average value, the average math growth percentile was 44.04. Time and partnership status were not statistically significant predictors of math growth percentiles. Statistically significant district-level predictors of average math growth percentiles included the proportion of non-resident students arriving to a district (positive) and district fund balance (negative). No district-level factors statistically significantly predicted math growth percentile change over time.

For the final model, across all included schools and years, with SQI scores and included school- and district-level factors at their average value, the average math growth percentile was 43.92. Time and partnership status were not statistically significant predictors of math growth percentiles. District-level factors were no longer statistically significant predictors of math growth percentiles. Patterns of the predictors were the same as the school-level factors model, except counselor-to-student ratios no longer statistically significantly predicted average math growth percentiles.

Discipline

Expulsions. For the RQ2 growth models of expulsions, indices of model fit only met thresholds of adequate fit for RMSEA (Table 24).

Table 24

RQ2 Expulsions Growth Models

<u>Model</u>	<u>N</u>	<u>Model Fit</u>				
		<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	333	29	$\chi^2(171) = 444.83^{***}$	0.75	0.75	0.07
District-level factors	400	27	$\chi^2(163) = 520.59^{***}$	0.74	0.74	0.07

Table 24 (cont'd)

Final Model	377	21	$\chi^2(139) = 412.17^{***}$	0.78	0.79	0.07
Parameter Estimates						
<i>Within Model</i>	<u>School Factors</u>		<u>District Factors</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
SQI (b_{iw1})	>0.01		-0.01*		>0.01	
Enrollment (b_{iw2})	>0.01		-		-	
Demographics (b_{iw3})	-0.01*		-		>0.01	
Mobility Rate (b_{iw4})	0.02**		-		0.01*	
School Staff (b_{iw5})	>0.01		-		-	
Support Staff (b_{iw6})	-0.03		-		-	
Counselors (b_{iw7})	0.01		-		-	
<u>Effects on Growth</u>						
SQI (b_{sw1})	>0.01		>0.01		>0.01	
Enrollment (b_{sw2})	>0.01		-		-	
Demographics (b_{sw3})	>0.01		-		>0.01	
Mobility Rate (b_{sw4})	>0.01		-		>0.01	
School Staff (b_{sw5})	>0.01		-		-	
Support Staff (b_{sw6})	>0.01		-		-	
Counselors (b_{sw7})	>0.01		-		-	
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	0.05***		0.06*		0.06*	
Slope Variance (σ_{sw}^2)	>0.01		>0.01		>0.01	
Covariance (cov_{isw})	>0.01		>0.01		>0.01	
<i>Between Model</i>	<u>School-Level</u>		<u>District-Level</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
Intercept (b_{ib})	0.22***		0.26***		0.25***	
Partnership Status (b_{ib1})	0.04		-0.06		-0.01	
Residents Out (b_{ib2})	-		-0.24		-	
Non-Residents In (b_{ib3})	-		0.31		-	
Expenditures (b_{ib4})	-		0.02		-	
Revenue (b_{ib5})	-		-0.01		-	
Fund Balance (b_{ib6})	-		>0.01		-	
<u>Effects on Growth</u>						
Slope (b_{sb})	>0.01		>0.01		>0.01	
Partnership Status (b_{sb1})	-0.01		-0.02**		-0.01	
Residents Out (b_{sb2})	-		0.03		-	
Non-Residents In (b_{sb3})	-		>0.01		-	
Expenditures (b_{sb4})	-		>0.01		-	
Revenue (b_{sb5})	-		>0.01		-	
Fund Balance (b_{sb6})	-		>0.01		-	

Table 24 (cont'd)

<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	0.06	0.03*	0.07
Slope Variance (σ_{sb}^2)	0.01*	>0.01	>0.01
Covariance (cov _{isb})	-0.01*	>0.01	-0.01*

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

For all RQ2 expulsions models, the within-level intercept variance was statistically significant, indicating there were changes over time in the proportion of students who received expulsions within schools. The between-level slope variance was statistically significant in the school-level factors model, indicating—in this model—there were changes over time in the proportion of students who received expulsions between districts. The between-level intercept variance was statistically significant in the district-level factors model, indicating—in this model—there were changes over time in the proportion of students who received expulsions within districts. The between-level slope and intercept covariance was statistically significant in the school-level factors and final models, indicating—in these models—a district's average expulsions was negatively associated with estimated change over time (i.e., lower expulsions associated with higher rates of change).

For the school-level factors model, across all included schools and years, with SQI scores and school-level factors at their average value, the average proportion of students who received expulsions was 0.22%. Time and partnership status were not statistically significant predictors of expulsions. Statistically significant school-level predictors of average expulsions included mobility rate (positive) and the latent factor of demographics representing the proportion of historically marginalized students (negative). None of the school-level factors statistically significantly predicted change over time.

For the district-level factors model, across all included schools and years, with SQI scores and district-level factors at their average value, the average proportion of students who received expulsions was 0.26%. Partnership status statistically significantly predicted change over time, such that partnership districts saw less growth in the average proportion of students who received expulsions. Time and district-level factors did not statistically significant predict expulsions.

Because none of the district-level factors were statistically significant, the final model includes only school-level factors. Across all included schools and years, with SQI scores and included school-level factors at their average value, the average proportion of students who received expulsions was 0.25%. Time and partnership status were not statistically significant predictors of expulsions. Only mobility rate remained a statistically significant (positive) predictor of average expulsions.

In-School Suspensions. For the RQ2 growth models of in-school suspensions, indices of model fit only met threshold of adequate fit for RMSEA (Table 25).

Table 25

RQ2 In-School Suspensions Growth Models

Model Fit						
Model	N	Parameters	Chi-Square	CFI	TLI	RMSEA
School-level factors	333	25	$\chi^2(50) = 105.63^{***}$	0.84	0.80	0.06
District-level factors	392	23	$\chi^2(47) = 129.52^{***}$	0.80	0.77	0.07
Final Model	341	21	$\chi^2(44) = 103.72^{***}$	0.83	0.81	0.06
Parameter Estimates						
<i>Within Model</i> Effects on Mean	School Factors		District Factors		Final Model	
SQI (b_{iw1})	-0.08*		-0.07		-0.07	
Enrollment (b_{iw2})	0.01		-		>0.01	
Demographics (b_{iw3})	>0.01		-		-	
Mobility Rate (b_{iw4})	>0.01		-		-	
School Staff (b_{iw5})	-0.09		-		-	
Support Staff (b_{iw6})	-0.12		-		-	

Table 25 (cont'd)

Counselors (b_{iw7})	0.58***	-	0.58***
<u>Effects on Growth</u>			
SQI (b_{sw1})	>0.01	>0.01	>0.01
Enrollment (b_{sw2})	-0.01**	-	-0.01*
Demographics (b_{sw3})	>0.01	-	-
Mobility Rate (b_{sw4})	-0.02	-	-
School Staff (b_{sw5})	-0.02	-	-
Support Staff (b_{sw6})	-0.02	-	-
Counselors (b_{sw7})	-0.01	-	-0.01
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	32.23	32.15	32.07
Slope Variance (σ_{sw}^2)	0.42	0.35	0.42
Covariance (cov_{isw})	-3.33	-3.07	-3.29
<u>Between Model</u>	<u>School-Level</u>	<u>District-Level</u>	<u>Final Model</u>
<u>Effects on Mean</u>			
Intercept (b_{ib})	6.01***	6.67***	6.15***
Partnership Status (b_{ib1})	-2.04	-2.31*	-0.47
Residents Out (b_{ib2})	-	-8.04*	-5.44
Non-Residents In (b_{ib3})	-	31.84**	16.00
Expenditures (b_{ib4})	-	0.26	-
Revenue (b_{ib5})	-	-0.08	-
Fund Balance (b_{ib6})	-	-0.01	-
<u>Effects on Growth</u>			
Slope (b_{sb})	-0.32	-0.42***	-0.34
Partnership Status (b_{sb1})	0.22	0.18	0.12
Residents Out (b_{sb2})	-	0.88	0.28
Non-Residents In (b_{sb3})	-	-5.08***	-2.87
Expenditures (b_{sb4})	-	-0.07	-
Revenue (b_{sb5})	-	0.05	-
Fund Balance (b_{sb6})	-	0.01	-
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	43.01***	6.81*	29.86*
Slope Variance (σ_{sb}^2)	0.48**	0.16**	0.32
Covariance (cov_{isb})	-3.81**	-0.52	-2.29

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

None of the within-level random effects were statistically significant, indicating little change over time in in-school suspensions among schools. For all models, the between-level

intercept was statistically significant, indicating there were changes over time in the proportion of students who received in-school suspensions within districts. For the school- and district-level factors models, the between-level slope was statistically significant, indicating—in these models—there were changes over time in the proportion of students who received in-school suspensions between districts. For the school-level factors model, the between-level slope and intercept covariance were also statistically significant, indicating—in this model—a district’s average in-school suspensions were negatively associated with estimated change over time (i.e., lower proportions of students suspended associated with higher rates of change).

For the school-level factors model, time and partnership status were not statistically significant predictors of in-school suspensions. Across all included schools and years with SQI scores and school-level factors at their average value, the average proportion of students who received in-school suspensions was 6.01%. Statistically significant school-level predictors of average in-school suspensions included SQI scores (negative) and counselor-to-student ratios (positive). Statistically significant school-level predictors of in-school-suspensions change over time included enrollment change (negative).

For the district-level factors model, time and partnership status were statistically significant and negative. Across all included schools and years, with SQI scores and district-level factors at their average value, the average proportion of students who received in-school suspensions was 6.67%, and suspensions were estimated to decrease by 0.42% each year. Partnership status was associated with a 2.31% decrease in the average proportion of students who received in-school suspensions. Statistically significant district-level predictors of average in-school suspensions included the proportion of resident students leaving a district (negative) and the proportion of non-resident students arriving to a district (positive). Statistically

significant district-level predictors of in-school suspensions change over time included the proportion of non-resident students arriving to a district (negative).

For the final model, time and partnership status were not statistically significant predictors of in-school suspensions. Across all included schools and years, with SQI scores and included school-level factors at their average value, the average proportion of students who received in-school suspensions was 6.15%. The patterns of the school-level factors, except SQI scores, remained the same, whereas district-level factors no longer predicted in-school suspensions.

Out-of-School Suspensions. For the RQ2 growth models of out-of-school suspensions, none of the indices of model fit met thresholds of adequate fit (Table 26). As such, I recommend interpreting the parameter estimates with caution.

Table 26

RQ2 Out-of-School Suspensions Growth Models

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
School-level factors	333	26	$\chi^2(49) = 170.88^{***}$	0.82	0.78	0.09
District-level factors	392	24	$\chi^2(46) = 222.64^{***}$	0.83	0.79	0.10
Final Model	392	20	$\chi^2(40) = 257.65^{***}$	0.72	0.68	0.12
Parameter Estimates						
<u>Within Model</u>	<u>School Factors</u>		<u>District Factors</u>		<u>Final Model</u>	
<u>Effects on Mean</u>						
SQI (b_{iw1})	-0.21**		-0.31***		-0.31***	
Enrollment (b_{iw2})	>0.01		-		-	
Demographics (b_{iw3})	0.02		-		-	
Mobility Rate (b_{iw4})	0.20		-		-	
School Staff (b_{iw5})	-0.32		-		-	
Support Staff (b_{iw6})	0.16		-		-	
Counselors (b_{iw7})	0.50		-		-	
<u>Effects on Growth</u>						
SQI (b_{sw1})	-0.01		>0.01		>0.01	
Enrollment (b_{sw2})	>0.01		-		-	
Demographics (b_{sw3})	>0.01		-		-	

Table 26 (cont'd)

Mobility Rate (b_{sw4})	>0.01	-	-
School Staff (b_{sw5})	>0.01	-	-
Support Staff (b_{sw6})	0.18	-	-
Counselors (b_{sw7})	-0.04	-	-
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	86.29***	111.61***	115.72***
Slope Variance (σ_{sw}^2)	1.69*	2.11**	2.15***
Covariance (cov_{isw})	-7.46**	-9.03**	2.15**
<u>Between Model</u>	<u>School-Level</u>	<u>District-Level</u>	<u>Final Model</u>
<u>Effects on Mean</u>			
Intercept (b_{ib})	13.72***	15.18***	14.29***
Partnership Status (b_{ib1})	0.53	2.34	3.05
Residents Out (b_{ib2})	-	-15.79***	-10.42*
Non-Residents In (b_{ib3})	-	17.34	14.52
Expenditures (b_{ib4})	-	-0.53	-
Revenue (b_{ib5})	-	0.66	-
Fund Balance (b_{ib6})	-	-0.23*	-0.20
<u>Effects on Growth</u>			
Slope (b_{sb})	0.48	0.27*	0.33
Partnership Status (b_{sb1})	-0.06	-0.56***	-0.58
Residents Out (b_{sb2})	-	3.93***	3.27**
Non-Residents In (b_{sb3})	-	-4.03*	-2.73
Expenditures (b_{sb4})	-	-0.05	-
Revenue (b_{sb5})	-	0.03	-
Fund Balance (b_{sb6})	-	0.08**	0.07*
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	23.76	2.79	5.25
Slope Variance (σ_{sb}^2)	0.97	0.02	0.03
Covariance (cov_{isb})	-4.26	-0.22	-0.25

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

For all models, the within-level intercept, slope, and covariance between the two were statistically significant, indicating there were changes over time in the proportion of students who received out-of-school suspensions within and between schools, and a school's average out-of-school suspensions was negatively associated with estimated change over time (i.e., lower proportion of students suspended associated with higher rates of change). None of the between-

level random effects were statistically significant, indicating little variation in out-of-school suspensions among districts over time.

For the school-level factors model, the effect of time and partnership status were not statistically significant predictors of out-of-school suspensions. Across all included schools and years, with SQI scores and school-level factors at their average value, the average proportion of students who received out-of-school suspensions was 13.72%. Only SQI scores emerged as a statistically significant predictor, negatively predicting average out-of-school suspensions.

For the district-level factors model, the effect of time and partnership status were statistically significant predictors of out-of-school suspensions. Across all included schools and years, with SQI scores and district-level factors at their average value, the average proportion of students who received out-of-school suspensions was 15.18%, and suspensions were estimated to increase by 0.27% each year. Partnership status was associated with a 0.56% decrease in out-of-school suspensions change over time. Statistically significant district-level predictors of average out-of-school suspensions included the proportion of resident students leaving a district (negative) and district fund balance (negative). Statistically significant district-level predictors of out-of-school suspensions change over time included the proportion of resident students leaving a district (positive), the proportion of non-resident students arriving to a district (negative) and district fund balance (positive).

Because none of the school-level factors were statistically significant, the final model includes only the significant district-level factors and SQI scores. The effect of time and partnership status were not statistically significant predictors of out-of-school suspensions. Across all included schools and years, with SQI scores and included district-level factors at their average value, the average proportion of students who received out-of-school suspensions was

14.29%. Statistically significant district-level predictors of average out-of-school suspensions included the proportion of resident students leaving a district (negative). Statistically significant district-level predictors of out-of-school suspensions change over time included the proportion of resident students leaving a district (positive) and district fund balance (positive).

Summary of RQ2 Findings

Overall, I encountered issues with model fit due to the large number of parameters relative to the cluster sample size. In addition to running these models sequentially (i.e., first identifying significant school-level factors, then district-level factors), I ran one large model with all school- and district-level factors included as a sensitivity analysis. Parameter estimates were largely similar in size, direction, and significance. In all instances except one, the only changes in statistical significance between the models were additional predictors emerging as statistically significant in the large models, suggesting my approach may have been more conservative. In the full model of ELA scores, mobility rate became non-significant, although the size and direction of the estimate were similar to the full model reported in Table 20.

Several structural factors emerged as statistically significant predictors of average student outcomes or their change over time in the direction I expected: mobility rate (negatively predicted attendance rates, ELA scores, math scores, and math growth percentiles; positively predicted expulsions), enrollment change (positively predicted attendance rate and ELA growth percentiles; negatively predicted in-school suspensions), counselor-to-student ratios (positively predicted ELA scores and math scores), the latent factor of demographics representing the proportion of historically marginalized students (negatively predicted ELA scores and math scores), and the proportion of resident students leaving a district (positively predicted out-of-school suspensions change over time).

However, there were also several findings that require more nuanced interpretations: counselor-to-student ratios positively predicted average in-school suspensions and negatively predicted average math growth percentiles, attendance rate change over time, and ELA growth percentile change over time; mobility rate positively predicted ELA score changes over time; the latent factor of demographics representing the proportion of historically marginalized students positively predicted math score changes over time; school staff-to-student ratios negatively predicted average ELA growth; support staff-to-student ratios negatively predicted average math growth; the proportion of non-resident students arriving negatively predicted attendance rate change over time; the proportion of resident students leaving a district negatively predicted average out-of-school suspensions; and a district's fund balance positively predicted out-of-school suspensions change over time.

As an additional sensitivity analysis, I also reran every final model that included the ratio of counselors-to-students or the ratio of school staff-to-students with a dataset that did not include the extreme values of paraeducators or counselors (see pg. 52). In these analyses, no patterns regarding school staff-to-student ratios changed, whereas counselor-to-student ratios no longer statistically significantly predicted change over time in ELA growth scores or average in-school-suspensions.

Partnership status did not statistically significantly predict any student outcomes or their linear change over time in the final models that included SQI scores and significant school- or district-level structural factors. Taken together—and caveated with the limitations of the cluster sample size—these findings provide mixed evidence of the *direction* of the impact of structural factors on student outcomes, but support the assertion that structural factors predict student outcomes over and above SQI scores and partnership status.

RQ3: Are the patterns identified in RQ1 and RQ2 the same when examining outcomes for different groups of students?

I only report findings for math growth percentiles, as it was the only outcome where all models successfully ran for each student subgroup. For clarity, I present simplified tables that include the unconditional growth model and final two models for RQ1 (the multilevel growth model including partnership status) and RQ2 (the multilevel growth model including significant school- and district-level predictors) for math growth percentiles by student subgroup.

Black Students

The residual variances for math growth percentiles for Black students at time points six and seven were too different in size compared to the remaining time points. Following established guidelines (Muthén & Muthén, 2021), I freely estimated the within-level residual variance at these three time points and constrained the remaining time points to be equivalent.

For RQ3 math growth percentile models for Black students, all indices of model fit met or exceeded all thresholds of adequate fit (Table 27).

Table 27

RQ3 Math Growth Percentile Growth Models for Black Students

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Unconditional	316	8	$\chi^2(6) = 12.06$	0.96	0.96	0.06
RQ1	316	13	$\chi^2(15) = 22.22$	0.97	0.96	0.04
RQ2	304	17	$\chi^2(19) = 26.95$	0.98	0.97	0.04
Parameter Estimates						
<i>Within Model</i>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})	40.92***		-	-		
SQI (b_{iw1})	-		-	0.19***		
School Staff (b_{iw2})	-		-	-0.37***		
<u>Effects on Growth</u>						
Slope (b_{sw})	0.81***		-	-		

Table 27 (cont'd)

SQI (b_{sw1})	-	-	0.01
School Staff (b_{sw2})	-	-	0.16**
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	43.83***	34.22***	24.62***
Slope Variance (σ_{sw}^2)	4.79***	4.18***	4.01***
Covariance (cov_{isw})	-7.12*	-5.61*	-6.34**
<u>Between Model</u>			
<u>Effects on Mean</u>			
Intercept (b_{ib})	-	42.08***	40.99***
Partnership Status (b_{ib1})	-	-2.71***	-0.55
<u>Effects on Growth</u>			
Slope (b_{sb})	-	0.31	0.40
Partnership Status (b_{sb1})	-	0.39	0.36
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	5.74**	1.84
Slope Variance (σ_{sb}^2)	-	0.61*	0.47
Covariance (cov_{isb})	-	-0.94	-0.16

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

For all growth models, the within-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating there were changes over time in math growth percentiles for Black students within and between schools, and a school's average math growth percentile for Black students was associated with its estimated change over time (i.e., lower average rates associated with higher estimated rates of change). For the RQ2 multilevel growth model, the between-level intercept and slope variance were also statistically significant, indicating—for this model—there were changes over time in Black students' math growth percentiles within and between districts.

For the unconditional model, the effect of time was statistically significant. Across all included schools and years, the average math growth percentile for Black students was 40.92, and growth percentiles were estimated to increase by 0.81 each year.

For the multilevel growth model that included partnership status, the effect of time was not statistically significant. Across all included schools and years and accounting for clustering within districts and partnership status, the average math growth percentile for Black students was 42.08. Partnership status negatively predicted the intercept, such that partnership status was associated with a 2.71 decrease in average math growth percentiles.

For the multilevel growth model that included statistically significant school- and district-level factors, the effect of time was not statistically significant. Only one factor emerged as significant in the school- and district-level factor models: school staff-to-student ratios. Across all included schools and years with SQI scores and school staff-to-student ratios at their average value, the average math growth percentile for Black students was 40.99. Partnership status was no longer a statistically significant predictor. School staff-to-student ratios negatively predicted average math growth percentiles for Black students, but positively predicted their change over time.

Latine Students

For the RQ3 math growth percentile models for Latine students, none of the indices of model fit met thresholds of adequate fit (Table 28). As such, I recommend interpreting parameter estimates with caution.

Table 28

RQ3 Math Growth Percentile Growth Models for Latine Students

<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Model Fit</u>			
			<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Unconditional	183	6	$\chi^2(8) = 21.54^{**}$	0.81	0.86	0.10
RQ1	183	11	$\chi^2(17) = 44.68^{***}$	0.71	0.73	0.09
RQ2	159	19	$\chi^2(25) = 62.24^{***}$	0.75	0.68	0.10

Table 28 (cont'd)

Parameter Estimates			
<i>Within Model</i>	<u>Unconditional</u>	<u>RQ1</u>	<u>RQ2</u>
<u>Effects on Mean</u>			
Intercept (b_{iw})	46.09***	-	-
SQI (b_{iw1})	-	-	0.14**
Enrollment (b_{iw2})	-	-	>0.01
Counselors (b_{iw3})	-	-	0.48***
<u>Effects on Growth</u>			
Slope (b_{sw})	0.31	-	-
SQI (b_{sw1})	-	-	0.02
Enrollment (b_{sw2})	-	-	>0.01
Counselors (b_{sw3})	-	-	-0.13*
<u>Random Effects</u>			
Intercept Variance (σ_{iw}^2)	44.07***	36.52***	29.60***
Slope Variance (σ_{sw}^2)	4.00	2.98	3.06
Covariance (cov_{isw})	-6.23	-4.26	-4.44
<i>Between Model</i>	<u>Unconditional</u>	<u>RQ1</u>	<u>RQ2</u>
<u>Effects on Mean</u>			
Intercept (b_{ib})	-	46.15***	45.85***
Partnership Status (b_{ib1})	-	-1.44	-0.85
Fund Balance (b_{ib2})	-	-	-0.09**
<u>Effects on Growth</u>			
Slope (b_{sb})	-	0.24	0.36
Partnership Status (b_{sb1})	-	0.55	0.90*
Fund Balance (b_{sb2})	-	-	0.01
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	6.48	2.34
Slope Variance (σ_{sb}^2)	-	0.74	0.67
Covariance (cov_{isb})	-	-1.34	-0.48

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

For all growth models, the within-level intercept variance was statistically significant, indicating there were changes over time in math growth percentiles for Latine students within schools. None of the between-level random effects were statistically significant, indicating little variation in math growth percentiles for Latine students among districts over time.

For the unconditional model, the effect of time was not statistically significant. Across all included schools and years, the average math growth percentile for Latine students was 46.09.

For the multilevel growth model that included partnership status, the effect of time was not statistically significant. After accounting for clustering within districts and partnership status, the average math growth percentile for Latine students was 46.15. Partnership status did not statistically significantly predict average growth percentiles for Latine students or their estimated change over time.

For the multilevel growth model that included statistically significant school- and district-level factors, the effect of time was not statistically significant. Three factors emerged as statistically significant in the school- and district-level factor models: enrollment change, counselor-to-student ratios, and dollars per student remaining in a district's fund balance. However, enrollment change was no longer statistically significant in the final model. Across all included schools and years with SQI scores and school- and district-level factors at their average value, the average math growth percentile for Latine students was 45.85. Partnership status did statistically significantly predict growth percentile change over time, such that partnership district schools were estimated to have more growth in math growth percentiles for Latine students. Counselor-to-student ratios positively predicted average math growth percentiles for Latine students, but negatively predicted their change over time. District fund balance negatively predicted average math growth percentiles for Latine students.

White Students

For RQ3 math growth percentile models for White students, most indices of model fit met or exceeded thresholds of adequate fit (Table 29).

Table 29*RQ3 Math Growth Percentile Growth Models for White Students*

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Unconditional	262	6	$\chi^2(8) = 13.58$	0.96	0.97	0.05
RQ1	262	11	$\chi^2(17) = 28.07^*$	0.94	0.94	0.05
RQ2	256	15	$\chi^2(21) = 21.12$	0.96	0.96	0.05
Parameter Estimates						
<u>Within Model</u>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})	46.55***		-	-		
SQI (b_{iw1})	-		-	0.25***		
Counselors (b_{iw2})	-		-	0.45***		
<u>Effects on Growth</u>						
Slope (b_{sw})	0.50*		-	-		
SQI (b_{sw1})	-		-	-0.01		
Counselors (b_{sw2})	-		-	0.13***		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	58.37***		49.38***	24.91***		
Slope Variance (σ_{sw}^2)	3.13		1.75	1.34		
Covariance (cov_{isw})	-4.81		-3.60	-2.14		
<u>Between Model</u>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{ib})	-		46.28***	46.99***		
Partnership Status (b_{ib1})	-		-2.34***	0.16		
<u>Effects on Growth</u>						
Slope (b_{sb})	-		0.14	0.22		
Partnership Status (b_{sb1})	-		0.29	0.26		
<u>Random Effects</u>						
Intercept Variance (σ_{ib}^2)	-		2.70	5.89		
Slope Variance (σ_{sb}^2)	-		1.47	1.47		
Covariance (cov_{isb})	-		-0.60	-1.61		

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

For all growth models, the within-level intercept variance was statistically significant, indicating there were changes over time in math growth percentiles for White students within schools. None of the other within-level or between-level random effects were statistically

significant, indicating little variation in math growth percentiles for White students among schools and districts over time.

For the unconditional model, the effect of time was statistically significant. Across all included schools and years, the average math growth percentile for White students was 46.55, and growth percentiles were estimated to increase by 0.50 each year.

For the multilevel growth model that included partnership status, the effect of time was not statistically significant. Across all included schools and years and accounting for clustering within districts and partnership status, the average math growth percentile for White students was 46.28. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 2.34 decrease in average math growth percentiles.

For the multilevel growth model that included statistically significant school- and district-level factors, the effect of time was not statistically significant. Only one factor emerged as statistically significant in the school- and district-level factor models: counselor-to-student ratios. Across all included schools and years, with SQI scores and counselor-to-student ratios at their average value, the average math growth percentile for White students was 46.99. Partnership status was no longer a statistically significant predictor. Counselor-to-student ratios positively predicted average math growth percentiles and their change over time for White students.

Students who Qualify for FRL

For RQ3 math growth percentile models for students who qualify for FRL, most indices of model fit met or exceeded thresholds of adequate fit (Table 30).

Table 30*RQ3 Math Growth Percentile Growth Models for Students who Qualify for FRL*

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Unconditional	375	6	$\chi^2(8) = 26.34^{***}$	0.95	0.96	0.08
RQ1	375	11	$\chi^2(17) = 56.09^{***}$	0.91	0.92	0.08
RQ2	364	15	$\chi^2(21) = 53.87^{***}$	0.95	0.94	0.07
Parameter Estimates						
<u>Within Model</u>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})	42.79***		-	-		
SQI (b_{iw1})	-		-	0.22***		
Counselors (b_{iw2})	-		-	0.29		
<u>Effects on Growth</u>						
Slope (b_{sw})	0.75***		-	-		
SQI (b_{sw1})	-		-	0.01		
Counselors (b_{sw2})	-		-	-0.10***		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	56.60***		40.02***	23.55***		
Slope Variance (σ_{sw}^2)	3.72***		2.14*	2.11*		
Covariance (cov_{isw})	-5.81**		-2.84	-3.17*		
<u>Between Model</u>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{ib})	-		43.93***	43.40***		
Partnership Status (b_{ib1})	-		-2.61**	-0.23		
<u>Effects on Growth</u>						
Slope (b_{sb})	-		0.04	0.08		
Partnership Status (b_{sb1})	-		0.15	0.24		
<u>Random Effects</u>						
Intercept Variance (σ_{ib}^2)	-		6.73**	5.26*		
Slope Variance (σ_{sb}^2)	-		1.51	1.41*		
Covariance (cov_{isb})	-		-1.72*	-1.87*		

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

For all models, the within-level intercept and slope variance were statistically significant, indicating there were changes over time in students who qualify for FRLs' math growth percentiles within and between schools. For the unconditional and RQ2 models, the covariance

between the two was also statistically significant, indicating—in these models—a school’s average math growth percentile for students who qualify for FRL was negatively associated with change over time (i.e., lower average percentiles associated with higher rates of change). For both multilevel growth models of RQ1 and RQ2, the between-level intercept variance and covariance between the intercept and slope were statistically significant, indicating there were changes over time in students who qualify for FRLs’ math growth percentiles within districts, and a district’s average growth percentiles were negatively associated with change over time. For the RQ2 model, the between-level slope variance was also statistically significant, indicating—in this model—there were changes over time in students who qualify for FRLs’ math growth percentiles between districts.

For the unconditional model, the effect of time was statistically significant. Across all included schools and years, the average math growth percentile for students who qualify for FRL was 42.79, and growth percentiles were estimated to increase by 0.75 each year.

For the multilevel growth model that included partnership status, the effect of time was not statistically significant. Across all included schools and years and accounting for clustering within districts and partnership status, the average math growth percentile for students who qualify for FRL was 43.93. Partnership status statistically significantly predicted the intercept, such that partnership status was associated with a 2.61 decrease in average math growth percentiles for students who qualify for FRL.

For the multilevel growth model that included statistically significant school- and district-level factors, the effect of time was not statistically significant. Only one factor emerged as statistically significant in the school- and district-level factor models: counselor-to-student ratios. Across all included schools and years, with SQI scores and counselor-to-student ratios at their

average value, the average math growth percentile for students who qualify for FRL was 43.

Partnership status was no longer a statistically significant predictor. Counselor-to-student ratios negatively predicted math growth percentile change over time for students who qualify for FRL.

Students with an IEP

Indices of model fit for the first two RQ3 math growth percentile models for students with an IEP only met thresholds of adequate model fit for RMSEA, whereas model fit indices for the final model met or exceeded all thresholds of adequate fit (Table 31).

Table 31

RQ3 Math Growth Percentile Growth Models for Students with an IEP

Model Fit						
<u>Model</u>	<u>N</u>	<u>Parameters</u>	<u>Chi-Square</u>	<u>CFI</u>	<u>TLI</u>	<u>RMSEA</u>
Unconditional	312	7	$\chi^2(7) = 20.70^{**}$	0.82	0.84	0.08
RQ1	312	12	$\chi^2(16) = 48.88^{***}$	0.78	0.78	0.08
RQ2	305	22	$\chi^2(26) = 37.00$	0.97	0.96	0.04
Parameter Estimates						
<i>Within Model</i>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{iw})	41.27***		-	-		
SQI (b_{iw1})	-		-	0.25		
Counselors (b_{iw2})	-		-	0.69**		
<u>Effects on Growth</u>						
Slope (b_{sw})	0.53*		-	-		
SQI (b_{sw1})	-		-	-0.01		
Counselors (b_{sw2})	-		-	-0.16		
<u>Random Effects</u>						
Intercept Variance (σ_{iw}^2)	48.34***		35.63***	17.84		
Slope Variance (σ_{sw}^2)	4.14*		3.95***	3.20		
Covariance (cov_{isw})	-8.62*		-7.58***	-5.39		
<i>Between Model</i>	<u>Unconditional</u>		<u>RQ1</u>	<u>RQ2</u>		
<u>Effects on Mean</u>						
Intercept (b_{ib})	-		42.22***	41.51***		
Partnership Status (b_{ib1})	-		-2.57**	-0.74		
Residents Out (b_{ib2})	-		-	4.93		
Expenditures (b_{ib4})	-		-	-0.75		

Table 31 (cont'd)

Revenue (b_{ib5})	-	-	0.72
<u>Effects on Growth</u>			
Slope (b_{sb})	-	0.45	0.32
Partnership Status (b_{sb1})	-	-0.03	-0.02
Residents Out (b_{sb2})	-	-	-0.29
Expenditures (b_{sb4})	-	-	0.28
Revenue (b_{sb5})	-	-	-0.27
<u>Random Effects</u>			
Intercept Variance (σ_{ib}^2)	-	6.72	2.05
Slope Variance (σ_{sb}^2)	-	0.25	0.07
Covariance (cov_{isb})	-	-1.07	-0.04

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

For the unconditional and RQ1 models, the within-level intercept variance, slope variance, and covariance between the two were statistically significant, indicating—in these models—there were changes over time in math growth percentiles for students with an IEP within and between schools, and a school’s average math growth percentile for students with an IEP was negatively associated with change over time (i.e., lower average percentiles associated with higher rates of change). None of the between-level random effects were statistically significant, indicating little variation in math growth percentiles for students with an IEP among districts over time.

For the unconditional model, the effect of time was statistically significant. Across all included schools and years, the average math growth percentile for students with an IEP was 41.27, and growth percentiles were estimated to increase by 0.53 each year.

For the multilevel growth model that included partnership status, the effect of time was not statistically significant. Across all included schools and years and accounting for clustering within districts and partnership status, the average math growth percentile for students with an IEP was 42.22. Partnership status statistically significantly predicted the intercept, such that

partnership status was associated with a 2.57 decrease in average math growth percentiles for students with an IEP.

For the multilevel growth model that included statistically significant school- and district-level factors, the effect of time was not statistically significant. Four factors emerged as statistically significant in the school- and district-level factor models: counselor-to-student ratios, the proportion of resident students leaving the district, expenditures per student, and revenue per student. However, counselor-to-student ratios was the only factor that remained statistically significant in the final model. Across all included schools and years, with SQI scores and included school- and district-level factors at their average, the average math growth percentile for students with an IEP was 41.51. Partnership status was no longer a statistically significant predictor. Counselor-to-student ratios positively predicted average math growth percentiles for students with an IEP.

Summary of RQ3 Findings

I was only able to examine math growth percentiles for RQ3, as it was the only outcome where all models successfully ran for each student subgroup. Partnership status statistically significantly predicted average math growth percentiles (negatively) for all student subgroups except Latine students. Partnership status did not statistically significantly predict math growth percentile linear change over time for any student subgroups. These findings suggest partnership district schools had worse overall math growth outcomes for all student groups except Latine students when compared to matched comparison district schools, but partnership status was not associated with differing trajectories in math growth outcomes over time for different groups of students. It is important to note the model for Latine students had relatively poor model fit compared to the other student subgroup models, likely due to the smaller sample size.

For the models that included structural factors, in addition to running these models sequentially (i.e., first identifying significant school-level factors, then district-level factors), I ran one large model with all school- and district-level factors included as a sensitivity analysis. Parameter estimates were largely similar in size, direction, and significance. There were two instances where the statistical significance of the structural factors changed: school staff-to-student ratios no longer predicted average math growth of Black students and counselor-to-student ratios no longer predicted math growth change over time of Latine students, although the size and direction of the estimates remained similar.

Several structural factors emerged as statistically significant predictors of average student outcomes or their change over time: counselor-to-student ratios (positively predicted average growth percentiles for Latine, White students, and students with an IEP; negatively predicted linear change for Latine students and students who qualify for FRL, positively predicted linear change for White students), school staff-to-student ratios (negatively predicted average growth percentiles for Black students; positively predicted linear change for Black students), and district fund balance (negatively predicted average growth percentiles for Latine students).

As an additional sensitivity analysis, I also reran every final model that included the ratio of counselors-to-students or the ratio of school staff-to-students with a dataset that did not include the extreme values of paraeducators or counselors (see pg. 52). In these analyses, no patterns regarding school staff-to-student ratios changed. Counselor-to-student ratios no longer statistically significantly predicted average math growth percentiles or their change over time for White students, Latine students, or students with an IEP, and negatively predicted average math growth percentiles for students who qualify for FRL, rather than their change over time.

Partnership status no longer statistically significantly predicted math growth percentiles or their linear change over time in the final models that included SQI scores and significant school- or district-level structural factors. Few structural factors emerged as statistically significant except counselor-to-student ratios. Taken together with the findings from the sensitivity analyses, these results suggest few of the structural factors included in my study impacted math growth percentiles for student subgroups, and those that did varied across subgroups.

DISCUSSION

Michigan's school accountability plan under ESSA was approved in November of 2017 (MDE, 2017, 2018a). Under the new accountability system, every three years MDE identifies public school districts with schools that are not meeting accountability standards and extends an invitation to enter into a District Partnership Agreement (Strunk et al., 2019). In this exploratory study, I used existing school- and district-level data from MI School Data and CRDC to examine three primary research questions. First, whether partnership district schools differed from matched comparison district schools on student outcomes over time. Second, whether proxies for structural factors (e.g., enrollment, financial status, resources) predicted student outcomes. Third, whether partnership district schools differed from matched comparison district schools in terms of equitable student outcomes.

My results suggest that changes in academic outcomes over time did not differ between partnership and matched comparison district schools. I also identified several structural factors that predicted student outcomes. Finally, in terms of equity, matched comparison district schools tended to have better average math growth percentile outcomes for all student subgroups, but few school and district structural factors emerged as statistically significant to explain these underlying differences. In my ensuing discussion, I address results relevant to each of my guiding questions in a dedicated subsection, and then elaborate on their implications for research, practice, and policy.

RQ1: How do partnership district schools—relative to their matched comparisons—experience changes in student outcomes over time?

My hypothesis that—over time—partnership district schools will experience more declines in student outcomes relative to matched comparison district schools was not supported

by my results. Partnership status did not statistically significantly predict linear changes over time for any of the student outcomes. Partnership status statistically significantly predicted average assessment scores, assessment growth, and out-of-school suspensions, such that partnership district schools had worse overall averages of these outcomes. Most outcomes were stable over time. Only ELA and math M-STEP scores and in-school suspension rates decreased over time, and their change over time did not differ between partnership and matched comparison district schools. I organize my discussion of my first research question with three sets of findings: (1) stability of student outcomes over time, (2) differences in demographic characteristics of partnership districts and matched comparison districts, and (3) student outcomes not predicted by partnership status.

Stability of Student Outcomes Over Time

My results for RQ1 suggest that some of the student outcomes of interest in MDE's school accountability system differ between partnership district schools and matched comparison district schools, and that these differences have been stable for a long time, suggesting the barriers partnership districts and their schools face are deeply rooted. Smaller, tailored programs tend to be ineffective at improving a broad set of school-wide outcomes; comprehensive school reform efforts (i.e., evidence-based programs or initiatives that focus on whole-school change; Borman et al., 2003) may be most effective for these schools (Rowan et al., 2004).

The partnership district model assumes that additional resources will enable schools and districts to start improving student achievement outcomes within 18 months, and to meet state standards within three years (MDE, 2017; Strunk et al., 2019). Meta analyses and systematic reviews of comprehensive school reform models suggest a minimum of three years of implementation is necessary to see any effects on academic outcomes, and five-to-ten years to

see strong effects (Borman et al., 2003; Maier et al., 2017), although there are some examples of faster effects (Johnston et al., 2020). It is unclear whether the supports provided by the partnership agreements meet the criteria for—or provide sufficient resources for schools to undergo—comprehensive school reform, which should include staff training and professional development, technical assistance, evaluation, parental involvement, and more (Borman et al., 2003). However, if the partnership model intends to provide this level of support, new parameters for the expected timelines to document improvements in student academic outcomes might be warranted. Additionally, for the partnership model to be most effective, continued support beyond the three years will be important as many school reform and improvement efforts are not sustained (Datnow, 2005).

Even if the partnership model provides adequate supports for schools to improve student achievement outcomes within three years (which is currently unknown due to the impact of the coronavirus pandemic; Strunk et al., 2021, 2022), it also outlines severe consequences—including replacing all school staff, having an Intermediate School District or the state take control, or closing the school—if the goals in the agreement are not met (Strunk et al., 2019).¹⁵ One of the concerns with the lack of federal oversight of ESSA is that states could easily maintain more traditional, punitive systems of school accountability—such as those under NCLB—rather than transforming their system to reflect ESSA’s new priorities of equity and school transformation (McGuinn, 2016). Further, MDE (2017) describes its ESSA plan as providing needed supports to schools rather than punishing them. In my study, student outcomes for partnership district schools did not change over time at a rate statistically significantly different than matched comparison district schools. A school accountability system that reflects

¹⁵ Partnership Agreements are publicly available at https://www.michigan.gov/mde/0,4615,7-140-81376_79956---.00.html

ESSA's goals of transformed, non-punitive school accountability should not penalize schools for being stable, rather, it should identify and provide the needed supports for schools to transform.

Overall, there is an opportunity to increase the Michigan school accountability system's alignment with ESSA to focus on providing evidence-based supports to these schools for sustained periods of time, such as comprehensive school reform, rather than sanctions for not meeting benchmarks.

Differences in Demographic Characteristics of Partnership Districts and Matched Comparison Districts

In my study, I originally attempted to identify matched comparison districts on size, locale, and all student characteristics. After encountering barriers during my initial exploration of the data trying to match districts on student race and ethnicity, I attempted—and failed—to identify matched comparison districts based on the average proportion of students who qualified for FRL. I could not find districts with a similar prevalence of economic disadvantage in similar locales and of similar sizes with sufficiently high SQI scores. The resulting sample of matched comparison districts served a statistically significantly lower proportion of students who qualified for FRL and lower proportions of Black students than partnership districts. It is important to note that I matched districts based on demographics for the 2009-10 school year and MDE's ESSA plan was not in place until 2017. However, descriptive statistics of student demographics suggest they were largely stable from the 2009-10 school year to the 2018-19 school year (Table 3).

ESSA includes an explicit focus on equity in response to NCLB research documenting that economically disadvantaged schools serving high proportions of Students of Color were disproportionately and unjustly subject to federal sanctions (Cook-Harvey et al., 2016; Kim &

Sunderman, 2005; McGuinn, 2016). Further, in MDE's (2017) ESSA plan, MDE commits to accounting for and addressing differences across student subgroups. However, my findings suggest schools and districts serving higher concentrations of Students of Color and students who qualify for FRL may be more likely to be the focus of intensive accountability efforts in Michigan. Additional research is warranted on the extent to which MDE's school accountability system is flagging schools with higher proportions of historically marginalized students. Researchers and practitioners in Michigan could explore the extent to which being flagged by accountability systems is associated with school-level demographic characteristics across all schools using SQI scores as the independent variable representing school accountability, rather than partnership status, to examine these patterns across all Michigan public schools.

Overall, there are additional opportunities for MDE to be more closely aligned to ESSA's goals of an equitable school accountability system, such as identifying and addressing elements of the school accountability system that are flagging schools serving high proportions of historically marginalized students.

Student Outcomes Not Predicted by Partnership Status

My results suggest the relationship between partnership status and disciplinary outcomes is unclear. Partnership status statistically significantly predicted higher out-of-school suspensions but not in-school suspensions or expulsions, and most models for disciplinary outcomes did not meet thresholds of adequate model fit. A goal in the shift from NCLB to ESSA was to promote holistic and well-rounded educational experiences by granting states flexibility in the indicators they include in their school quality measures (Bae, 2018; Cardichon & Darling-Hammond, 2017). Given that researchers have previously documented the persistent, negative impact of exclusionary discipline on student achievement and dropout (Noltemeyer et al., 2015;

Rumberger & Losen, 2016), it is important that researchers explore disciplinary and other non-academic outcomes in the school accountability context in additional samples, such as the extent to which they differ between schools rated as high performing and schools rated as low performing. These studies would further clarify the extent to which certain models of school accountability (e.g., the partnership agreement model) accurately identify schools and districts that need support providing positive, well-rounded educational experiences for students.

Similarly, attendance rate did not differ between partnership and matched comparison districts and most models did not meet thresholds of adequate model fit. Attendance rate positively impacts student achievement outcomes and is often a key component of school improvement efforts; higher attendance rates are associated with higher GPAs and standardized test scores (Gershenson et al., 2017; Gottfried, 2009, 2010). Michigan's school accountability system focuses on chronic absenteeism, rather than attendance rate (MDE, 2017). Researchers have suggested the detrimental impact of chronic absenteeism on individual student achievement might be the reason for its focus in ESSA (Patnode et al., 2018). However, I could not find published research that explores the difference between school-level chronic absenteeism and attendance rates' predictive ability of school-level student achievement. This is an opportunity for future research. If chronic absenteeism is simply a more discriminating variable than attendance rate (i.e., it differs more across schools) and does not better predict student outcomes, the merit of its focus in school accountability over attendance rate should be called into question.

Overall, my results suggest the relationship between attendance rate and disciplinary outcomes and partnership status is unclear. These outcomes should be considered in future research and continuous quality improvement efforts for Michigan's school accountability system.

RQ2: How do proxies for structural factors predict student outcomes?

My hypothesis that *measured indicators of structural barriers will predict worse student outcomes and indicators of structural resources will predict better student outcomes* was partially supported by my results. Some indicators of structural barriers negatively impacted average student academic outcomes (enrollment change, mobility rate, school demographics, and the proportion of resident students leaving), whereas other indicators of structural factors were related to student outcomes in unanticipated or inconsistent directions (counselor-to-student ratios, school staff-to-student ratios, support staff-to-student ratios, the proportion of non-resident students arriving, and district fund balance). Partnership status did not statistically significantly predict any student outcomes or their linear change over time in final models that included SQI scores and statistically significant structural factors. As such, my results suggest that—although structural factors predict student outcomes over and above SQI scores and partnership status—the strength and direction of their predictive associations varies across the structural factor or student outcome.

I used an ecological approach (Bronfenbrenner, 1977, 1979) to identify three categories of measured indicators of structural factors relevant to Michigan schools based on available data: (1) student enrollment characteristics (e.g., demographics); (2) student mobility characteristics (e.g., mobility rate); and (3) school and district resources (e.g., staff-to-student ratios). In my ensuing discussion of my second research question, I first review the overall findings and then findings for each of these structural factors.

Overall Findings

In my study, structural factors predicted student outcomes, even after accounting for SQI scores and partnership status. Critics of school accountability policies argue many of the

educational outcomes of interest in accountability systems are explained by the ecological contexts of students and schools, rather than the quality of educational experiences provided by schools (Schneider et al., 2021). The types of structural factors I included are either completely outside of a school or district's locus of control (e.g., student demographics) or heavily reliant on funding (e.g., staff-to-student ratios). This finding further supports the need for comprehensive school reform (Borman et al., 2003; Rowan et al., 2004) and intervening further out in a school's ecological system (e.g., macrosystem), such as providing district and community supports (Eccles & Roeser, 2008; Lewallen et al., 2015).

The indicators I included were either measured at the school (i.e., exosystem) or district (i.e., macrosystem) level but represent structural factors that span various levels of the ecological system. Individuals with additional access to student and community data could incorporate indicators of structural factors at different levels (e.g., socioeconomic status at the individual, school, and community level) and across different domains (e.g., health) to identify the best areas to provide supports. Additionally, these data can be used in future research to identify which student outcomes are most heavily influenced by structural factors that schools cannot change, which has important implications for their use in school accountability metrics (Schneider et al., 2021).

Converging with results from my first research question, the only outcomes that SQI scores did not statistically significantly predict were in-school suspensions and expulsions. An important research finding from NCLB was that if accountability systems did not include specific metrics on a topic or outcome, they were less likely to be emphasized—and in some instances lost funding—in school settings (Pederson, 2007). Michigan and many other states have yet to include indicators that reflect holistic aspects of students' educational experiences,

such as disciplinary practices (Dusenbury et al., 2018; MDE, 2017; Schneider et al., 2021). These findings highlight disciplinary outcomes as potential areas of interest for school accountability policies. Not only do these outcomes predict some of the achievement metrics of interest under ESSA (e.g., school dropout and test scores; Noltemeyer et al., 2015; Rumberger & Losen, 2016), they are aligned with ESSA's goals. ESSA requires states to develop and implement support plans to improve school climate and recommends these plans specify how schools will reduce exclusionary discipline practices (i.e., out-of-school suspension, expulsion), providing awards for districts or states that aim to transform school climate and discipline by shifting to focus on restorative practices (Adler-Greene, 2019; Bae, 2018; McGuinn, 2016). Further, from an ecological perspective, disciplinary policies and practices are within a school's locus of control.

My overall findings for RQ2 suggest that—under Michigan's new school accountability system—schools are being held accountable to metrics that may be influenced by factors largely outside of their control. To increase alignment with MDE's (2017) and ESSA's goals of an equitable, non-punitive system of school accountability, future research is warranted on the areas Michigan's accountability system could consider for school quality metrics, including the structural factors in my study and disciplinary or other non-academic outcomes. This research could focus on the identification of specific resources and supports that impact school outcomes in Michigan, which in turn could inform school accountability policy and practice.

Student Enrollment Characteristics

Regarding student enrollment characteristics, serving a higher proportion of students from historically marginalized backgrounds negatively predicted average ELA and math scores. These findings build upon the substantial body of evidence that school-level socioeconomic

status is negatively associated with achievement outcomes (Hegedus, 2018; Perry & McConney, 2010; van Ewijk & Slegers, 2010), Students of Color are more likely to attend schools with lower socioeconomic status (Hussar et al., 2020), and standardized assessment scores are strongly correlated with student demographics (Kane & Staiger, 2002; Schneider et al., 2021). However, student demographics did not predict assessment growth; in fact, the proportion of historically marginalized students positively predicted math score change over time. This finding—taken together with the RQ1 result that assessment scores were negatively declining over time whereas assessment growth remained stable—aligns with prior research documenting assessment scores and growth as distinct, unrelated measures and highlighting assessment growth as a better indicator for school accountability considerations, as it tends to be less correlated with student demographics (Downey et al., 2008; Hegedus, 2018). Further, the finding that schools serving students who face economic and social barriers are making gains toward improving math scores highlights the importance of longitudinal studies in school accountability research and education research, broadly.

Under ESSA, states are required to track and report outcomes for the general student population and for student subgroups (Darling-Hammond et al., 2016; McGuinn, 2016). These data should be used to inform school accountability policy and practice. If certain student subgroups are performing worse for the included outcomes across all Michigan schools, it could be disproportionately affecting the SQI scores of schools serving higher concentrations of these student subgroups, especially considering student subgroup performance is weighted less in these scores (MDE, 2017). Future studies could explore the standardized test score weights in MDE's SQI scores. For instance, researchers could use existing data to examine the extent to which changing the weight of schools' average standardized ELA and math scores relative to the

weight of student subgroups' ELA and math scores changes which schools—and what types of schools (e.g., demographics)—are flagged under the current accountability system.

Overall, serving higher proportions of historically marginalized students was associated with lower ELA and math scores but not growth, suggesting assessment growth might be an interesting area for equitable school accountability research and practice to explore. Further, school accountability policy makers in Michigan should consider the consistent, pervasive relationship between standardized assessment scores and student demographics (Kane & Staiger, 2002; Schneider et al., 2021) and explore opportunities within MDE's ESSA plan to promote equity.

Student Mobility Characteristics

In my study, different student mobility characteristics emerged as statistically significant predictors of student outcomes and appeared to differ between partnership and matched comparison district schools, suggesting student mobility and associated policies may be important to consider in school accountability. The direction of many of the relationships between student mobility characteristics and student outcomes reflected what I originally anticipated. For instance, enrollment loss negatively predicted average attendance rates and ELA growth percentiles, and mobility rate positively predicted expulsions and negatively predicted average attendance rates, ELA scores, math scores, and math growth percentiles. However, several relationships between student mobility characteristics and student outcomes were in unanticipated directions. For instance, the proportion of non-resident students arriving negatively predicted attendance rate change over time and mobility rate positively predicted ELA score changes over time.

School choice policies vary state-to-state, but they allow students to attend schools and districts outside their residential zone with the goal of incentivizing schools to improve their quality through competition (Logan, 2018). Previous research documents some of the potential consequences of school choice, such as negative academic outcomes for students that leave schools through voucher programs (Abdulkadiroglu et al., 2015; Carnoy, 2017) and declining enrollment (Arsen et al., 2019; Garnett, 2014). In Michigan, intermediate school districts decide whether to enroll in school choice policies that allow schools to enroll nonresident students without permission from the resident school or district (The State School Aid Act of 1979, 1979a; 1979b). Findings regarding student mobility were somewhat mixed, but overall seemed to suggest different characteristics of student mobility—including losing *or* gaining students—were negatively associated with student outcomes. These findings have important implications for the intersection of school choice and school accountability policies. Some states have school accountability policies that facilitate or expedite school choice approval for students attending schools not meeting their performance benchmarks (Bross et al., 2016; Carnoy, 2017; Feng et al., 2018). If states implement these policies, school accountability systems should consider student mobility in school quality measures to offset the potential negative impact these policies have on student outcomes.

Further, descriptive statistics from my study suggest that partnership district schools may have disproportionately experienced student mobility. Partnership districts had higher average enrollment loss ($M = -88.31$, $SD = 247.89$), mobility rate ($M = 19.20\%$, $SD = 16.19\%$), and proportion of resident students leaving ($M = 0.79\%$, $SD = 0.48\%$) relative to matched comparison districts (enrollment change $M = -20.59$, $SD = 148.24$; mobility rate $M = 11.50\%$; $SD = 14.00$; proportion of resident students leaving $M = 0.26\%$, $SD = 0.16\%$). This, coupled with the

finding that partnership districts served higher proportions of students from historically marginalized backgrounds, aligns with prior research documenting associations between school choice and the economic and racial segregation of schools (Glazerman & Dotter, 2017; Mordechay & Orfield, 2017; Roda & Wells, 2013; Winters, 2015).

Overall, my results regarding student mobility suggest there may be some negative associations between student mobility and academic outcomes. While future research is warranted regarding the strength and direction of the relationships between various student mobility characteristics and student outcomes over time, school choice and student mobility could be important factors to consider in school accountability systems.

School and District Resources

Regarding school and district resources, only staff-to-student ratios emerged as statistically significant predictors of student outcomes, and primarily in unanticipated directions. Counselor-to-student ratios positively predicted average ELA scores and math scores, but also positively predicted average in-school suspensions and negatively predicted average math growth percentiles, attendance rate change over time, and ELA growth percentile change over time. School staff-to-student ratios and support staff-to-student ratios negatively predicted average ELA and math growth percentiles, respectively. Thus, overall findings of the direct relationships between school and support staff and school-level student outcomes in this sample are unclear.

Guided by an ecological framework, I conceptualized school staff as an important school resource, in line with prior research associating higher staff-to-student ratios (i.e., fewer staff members per student) with negative student outcomes, such as bullying and dropping out of high school (Christle et al., 2007; Waasdorp et al., 2011) and lower staff-to-student ratios (i.e., more

staff members per student) with positive outcomes, such as student discipline and graduation rates (Lapan et al., 2012). I generated time invariant predictors (e.g., the average school staff-to-student ratio across all years of available data) rather than time varying predictors (e.g., the average school staff-to-student ratio each year) both due to the exploratory nature of my study and the limited sample size. Now that the foundation is set for which factors might predict student outcomes, future studies could examine these relationships over time as there may be short term effects of increasing certain supports. For instance, needing more school staff might be an indication that a school needs additional supports, thus being overall negatively associated with student outcomes. However, year-to-year differences in school staff ratios and their relationships with student outcomes might yield different patterns.

In most models, district resources did not statistically significantly predict student outcomes. The only statistically significant finding was not in a direction I anticipated: a higher fund balance was associated with greater increases in out-of-school suspension rates over time. This finding is difficult to interpret on its own. Across the years of my study, partnership districts—on average—earned and spent more dollars per student but had fewer dollars per student remaining in their fund balance compared to matched comparison districts (see Table 9 on pg. 64), meaning schools and districts rated as underperforming may be receiving more resources per student. This is a positive finding, as previous research indicates Michigan’s school funding system did not account for the costs of educating students facing additional barriers (Arsen et al., 2016; 2019). However, the financial variables in my models do not include other social and fiscal resources schools benefit from, such as private donations or fundraisers. Private donations to public schools have grown, concurrently with and in contrast to efforts to create a public school funding system less reliant on local property taxes (Frisch, 2017). These private

resources might be an important factor currently unaccounted for in school accountability and state data systems.

Overall, few school and district resources emerged as statistically significant predictors of student outcomes in my study, and largely not in the directions I originally anticipated. This is an important area of exploration for future research, such as the examination of time varying predictors, and school accountability policy and practice, such as accounting for private donations to schools.

RQ3: Are the patterns identified in RQ1 and RQ2 the same when outcomes for different groups of students?

I aimed to clarify competing hypotheses for my third research question: *(1) Matched comparison district schools have more equitable student outcomes (i.e., student outcome trends are positive and similar across subgroups), and these outcomes are predicted by structural factors; or (2) Partnership and matched comparison district schools do not differ in equitable student outcomes (i.e., student outcome trends are similar across subgroups for partnership and matched comparison district schools), and these outcomes are not predicted by structural factors.* Overall, my results provide limited support for the first assertion. I only report findings for math growth percentiles as it was the only outcome where all models successfully converged for each student subgroup. Partnership status statistically significantly and negatively predicted average math growth percentiles for all student subgroups except Latine students. Few structural factors emerged as statistically significant, making it difficult to disentangle the underlying reasons for these differences. It is also important to note that patterns from my RQ3 findings likely would not apply to other academic outcomes in this study; previous research implies assessment growth is less related to student demographic characteristics than assessment scores

(Hussar et al., 2020) and the proportion of historically marginalized students enrolled in a school was a statistically significant predictor of assessment scores in RQ2.

My findings imply that Michigan's school accountability system is predominantly flagging schools where math growth was worse for most students, not just students who were White, did not qualify for FRL, and did not have an IEP. However, the underlying structural factors that might enable matched comparison districts and schools to better support their students' math growth were unclear. Researchers have debated whether higher performing schools are better able to serve all students—including those historically disenfranchised by the education system—due to an increase in resources and supports, or whether they are categorized as higher performing because they tend to serve populations that are not historically disenfranchised by the education system (Chambers et al., 2014; Gaddis & Lauen, 2014; Harris & Herrington, 2006). Findings related to math growth in my sample partially suggest the former.

My RQ3 findings related to structural factors that statistically significantly predicted math growth percentiles across student subgroups differed from the structural factors that emerged as important for math growth percentiles across all students in RQ2 (mobility rate and support staff-to-student ratios). Initially, counselor-to-student ratios appeared to be an important, positive predictor for Latine students, White students, and students with an IEP. However, after I conducted a sensitivity analysis using a dataset without extreme values, counselor-to-student ratios no longer predicted any positive student outcomes across student subgroups. Further, the other structural factors that emerged as statistically significant were in unexpected or mixed directions: school staff-to-student ratios negatively predicted average math growth percentiles but positively predicted their linear change for Black students, and district fund balance negatively predicted average math growth percentiles for Latine students. Additionally, math

growth percentiles for Latine students did not differ between partnership and matched comparison district schools, whereas all other student subgroups in matched comparison district schools had higher math growth percentiles.

Together, these findings suggest different factors might matter more and in different ways for different students, aligning with prior research documenting different responses to school supports and practices by student subgroups (e.g., students with disabilities and Black students; Lane et al., 2007; Gregory et al., 2011; Pena-Shaff et al., 2019). There are promising interventions and practices that focus on providing different types of support to different student populations based on identified barriers or needs. For instance, students from low-income families may have less access to the resources that prepare them for college; providing follow-up guidance counseling for these students the summer after high school graduation increases rates of postsecondary attendance (Castleman et al., 2012). For Students of Color, racial and ethnic representation in their school setting impacts important achievement outcomes. Black and Latine students who attend schools with teachers or school leadership with the same racial and ethnic background have higher enrollment in gifted programs (Grissom et al., 2017). Having higher Black-to-White or Latine-to-White teacher ratios is associated with decreases in the suspension gaps between Black and White students and Latine and White students, respectively (Hughes et al., 2020). For Latine students, having teachers of the same ethnicity is associated with increased math and reading outcomes if the teaching staff is diverse (Banerjee, 2018).

Future research is warranted in this area, particularly given the small sample sizes across student subgroups in my study. Researchers could examine the relationships between structural factors and student subgroup outcomes while accounting for SQI scores rather than focusing on partnership status, allowing for a sample of all Michigan schools with SQI scores. Further,

researchers or analysts with access to student-level data could identify these relationships at the individual level, exploring interactions between individual student characteristics, their access to certain school supports (e.g., counselors, social workers), and their academic outcomes. These types of studies would have important implications for how MDE and other education systems can provide the most effective supports to promote student equity. Importantly, many scholars have highlighted the benefits of listening to the voices of youth with historically marginalized identities to understand the supports that will ensure these students can achieve their full potential (Davis et al., 2019; Huerta et al., 2020; Warren et al., 2016). Qualitative and participatory studies with youth to identify barriers and solutions to promoting equitable student outcomes should inform the policies and practices of schools whose student subgroups are categorized as not meeting performance standards.

Limitations

This study was exploratory in nature. I did not intend for my findings to generalize beyond Michigan due to the heterogeneity in education systems across states, although the patterns might inform research and practice for states with similar school accountability metrics and policies. There are several limitations to my study that arise from the nature of the data sources and the sample size. It was important for me to use the data as-is; these data are used by the Office of Civil Rights and MDE to inform research and policy. However, I did conduct several sensitivity analyses and I offer caveats for some of my findings related to specific variables where there may be data quality concerns.

Some of the data I sourced from CRDC had unanticipated, and potentially unrealistic, characteristics. For instance, I removed 94 suspension values that were greater than 100% (in-school suspensions $n = 8$; out-of-school suspensions $n = 86$). Even after removing these values,

out-of-school suspension rates appeared high (overall average of 14.54%), especially when compared to in-school suspension rates (overall average of 2.52%). CRDC highlights data quality as a potential issue on its website, as schools and districts are ultimately responsible for reporting the data and assuring its quality.¹⁶ It is possible that some schools reported the total number of suspensions, rather than the unique number of students receiving suspensions the variables intend to capture, inflating these rates. Additionally, due to a number of schools not having disciplinary data, there were instances where districts lacked within cluster variation (i.e., many schools were missing values within a given year).

Although I did not remove many counselor observations due to the number of counselors being greater than the number of students ($n = 2$), I noticed two observations that fell just under the implausible data thresholds for school staff-to-student ratios *and* appeared significantly higher than the second-highest observation (counselor-to-student ratio of 99/100, the next largest value was 4.5/100; instructional paraeducator-to-student ratio of 87/100, the next largest value was 45/100). As a sensitivity analysis, I reran any final models that included school staff or counselors as a statistically significant predictor with these outlying values removed. For my RQ3 analyses, which had substantially smaller sample sizes than my other models, counselors emerged as a statistically significant predictor of math growth percentiles for almost all student subgroups. When I conducted the sensitivity analysis, these relationships were no longer statistically significant.

Despite my best efforts to match districts on student economic status, the proportion of students qualifying for FRL was statistically significantly higher at baseline for partnership district schools ($M = 78.66\%$; $SD = 19.93\%$) compared to matched comparison district schools

¹⁶ <https://ocrdata.ed.gov/resources/datanotes>

($M = 59.33\%$; $SD = 16.88\%$). The adjusted difference between these two averages was large ($g = -1.08$, 95% CI = $[-1.27, -0.88]$). As described earlier in the discussion, this is a meaningful finding in and of itself, indicating schools in districts impacted by the partnership district model systematically differ from schools that are not on non-school quality-related factors. However, it also meant that my attempt to isolate certain student demographic characteristics (e.g., race and ethnicity) was not successful, although I accounted for demographic characteristics in my models using a latent factor.

The sampling frame of my study also affected my models. Because I focused on the 12 public school districts with partnership agreements, I ended up with 24 districts in my study and encountered convergence issues in my larger models where I estimated many different parameters (e.g., the models with school- and district-level factors). Similarly, I was unable to examine most student outcomes for my third research question due to small sample sizes and a lack of within cluster variation for some student subgroups. Future iterations of this study could explore these trends across the state of Michigan, using the entire sample reported in MI School Data, which would offer an increased sample size as well as be generalizable to Michigan schools. These iterations would likely need to focus on SQI scores alone, rather than partnership district status, to avoid the unequal sampling of partnership to non-partnership districts.

Finally, there are broad limitations to my use of measured indicators as proxies for structural factors. For instance, I conceptualized student demographics as a proxy for the underlying barriers students from historically marginalized identities overcome in educational settings. I acknowledge the limitations of this approach and its potential to reinforce narratives of the (under)performance of students from historically marginalized backgrounds. Only focusing on differences in achievement outcomes across student subgroups can detract from the structural

factors that underly these differences and hinders our ability to address educational inequities (Chambers, 2009; Gouvea, 2021; Milner IV, 2013). I attempted to include a variety of indicators reflecting structural factors outside of student demographics, and hope the findings of this exploratory study can serve as a starting point to understanding and accounting for the ecological contexts of students, schools, and districts in school accountability systems.

CONCLUSION

I had three overarching aims in this exploratory study: (1) examine the extent to which partnership district schools differed from matched comparison district schools on student outcomes over time; (2) examine the extent to which proxies for structural factors impacted student outcomes; and (3) examine the extent to which partnership district schools differed from matched comparison district schools in terms of equitable student outcomes.

For my first aim, I found that partnership status was statistically significantly and negatively associated with average student academic outcomes but not their change over time, whereas its association with student disciplinary outcomes and attendance was unclear. For my second aim, I found that some measured indicators suggested structural barriers negatively impacted average academic student outcomes, while others were related to student outcomes in unanticipated or inconsistent directions. For my third aim, I found partnership status statistically significantly and negatively predicted average math growth percentiles for all student subgroups except Latine students, and few structural factors emerged as statistically significant. Additionally, despite my best efforts to find a similar matched sample, partnership districts systematically differed from matched comparison districts on student characteristics, typically serving a higher proportion of historically marginalized students.

A key takeaway from this study is structural factors outside of the locus of control of individual schools or districts accounted for student outcomes over and above measures of school quality and partnership status. This supports the recommendation to abandon narratives that student achievement under the current school accountability policies is entirely the responsibility of the schools they attend, and focus on ongoing long-term supports for schools, districts, and communities. Overall, my findings suggest that in the early stages of the

implementation of its ESSA plan, there is room for growth towards achieving MDE's (2017) goals of promoting holistic, equitable educational outcomes.

REFERENCES

- Abdulkadiroglu, A., Pathak, P., & Walters, C. (2015). *Free to Choose: Can School Choice Reduce Student Achievement?* <https://doi.org/10.3386/w21839>
- Adler-Greene, L. (2019). Every Student Succeeds Act: Are schools making sure every student succeeds? *Touro Law Review*, 35(1), 11–23.
- Arsen, D., Delpier, T., & Nagel, J. (2019). *Michigan school finance at the crossroads: A quarter century of state control*.
- Arsen, D., Deluca, T., Ni, Y., & Bates, M. (2016). Why districts get into financial trouble and why: Michigan's story. *Journal of Education Finance*, 42(2), 100–126.
- Arsen, D., & Ni, Y. (2012). The effects of charter school competition on school district resource allocation. *Educational Administration Quarterly*, 48(1), 3–38. <https://doi.org/10.1177/0013161X11419654>
- Arsen, D., Plank, D., & Sykes, G. (1999). *School choice policies in Michigan: The rules matter*.
- Assari, S. (2019). Race, education attainment, and happiness in the United States. *International Journal of Epidemiologic Research*, 6(2), 76–82. <https://doi.org/10.15171/ijer.2019.14>
- Bae, S. (2018). Redesigning systems of school accountability: A multiple measures approach to accountability and support. *Education Policy Analysis Archives*, 26, 8. <https://doi.org/10.14507/epaa.26.2920>
- Bailey, M. J., & Dynarski, S. M. (2011). Inequality in postsecondary education. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 117–132). Russell Sage Foundation.
- Banerjee, N. (2018). Effects of teacher-student ethnoracial matching and overall teacher diversity in elementary schools on educational outcomes. *Journal of Research in Childhood Education*, 32(1), 94–118. <https://doi.org/10.1080/02568543.2017.1393032>
- Bauer, L., Liu, P., Schanzenbach, D. W., & Shambaugh, J. (2018). *Reducing Chronic Absenteeism under the Every Student Succeeds Act*.
- Bogart, W. T., & Cromwell, B. A. (2000). How Much Is a Neighborhood School Worth? *Journal of Urban Economics*, 47(2), 280–305. <https://doi.org/10.1006/juec.1999.2142>
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230. <https://doi.org/10.3102/00346543073002125>
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513–531.

- Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, 34, 844–850.
- Bross, W., Harris, D. N., Liu, L., Adhikari, B., Alm, J., Barreca, A., Lincove, J., Larsen, M., Teles, D., Zimmer, R., & Engberg, J. (2016). *The effects of performance-based school closure and charter takeover on student performance*.
- Brummet, Q. (2014). The effect of school closings on student achievement. *Journal of Public Economics*, 119, 108–124. <https://doi.org/10.1016/j.jpubeco.2014.06.010>
- Burchinal, M., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., McLoyd, V., & Pianta, R. (2011). Examining the Black-White achievement gap among low-income children using the NICHD Study of Early Child Care and Youth Development. *Child Development*, 82(5), 1404–1420. <https://doi.org/10.1111/j.1467-8624.2011.01620.x>
- Burdick-Will, J., Ludwig, J., Raudenbush, S. W., Sampson, R. J., Sanbonmatsu, L., Sharkey, P., Betts, J., Blank, R., Duncan, G., Katz, L., Kling, J., Murnane, R., & Rivkin, S. (2011). Converging evidence for neighborhood effects on children’s test scores: An experimental, quasi-experimental, and observational comparison. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children’s Life Chances* (pp. 255–276). Russell Sage Foundation.
- Cardichon, J., & Darling-Hammond, L. (2017). *Advancing Educational Equity for Underserved Youth How New State Accountability Systems Can Support School Inclusion and Student Success*. <https://learningpolicyinstitute.org/product/>
- Carnoy, M. (2017). *School vouchers are not a proven strategy for improving student achievement*.
- Castleman, B. L., Arnold, K., & Wartman, K. L. (2012). Stemming the tide of summer melt: An experimental study of the effects of post-high school summer intervention on low-income students’ college enrollment. *Journal of Research on Educational Effectiveness*, 5(1), 1–17. <https://doi.org/10.1080/19345747.2011.618214>
- Chambers, T. V. (2009). The “Receivment Gap”: School tracking policies and the fallacy of the “Achievement Gap.” *Journal of Negro Education*, 78(4), 417–431. <https://www.jstor.org/stable/25676096>
- Chambers, T. V., Huggins, K. S., Locke, L. A., & Fowler, R. M. (2014). Between a “ROC” and a school place: The role of Racial Opportunity Cost in the educational experiences of academically successful Students of Color. *Educational Studies*, 50(5), 464–497. <https://doi.org/10.1080/00131946.2014.943891>
- Christle, C. A., Jolivette, K., & Michael Nelson, C. (2007). School characteristics related to high school dropout rates. *Remedial and Special Education*, 28(6), 325–339. <https://doi.org/10.1177/07419325070280060201>
- Close, K., Amrein-Beardsley, A., & Collins, C. (2018). *State-level assessments and teacher*

- evaluation systems after the passage of the Every Student Succeeds Act: Some steps in the right direction. <http://nepc.colorado.edu/publication/state-assessment>
- Cook-Harvey, C. M., Darling-Hammond, L., Lam, L., Mercer, C., & Roc, M. (2016). *Equity and ESSA: Leveraging Educational Opportunity Through the Every Student Succeeds Act*.
- Dahlin, M., & Cronin, J. (2010). *Achievement Gaps and the Proficiency Trap*.
- Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. <https://edpolicy.stanford.edu>.
- Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly*, 41(1), 121–153. <https://doi.org/10.1177/0013161X04269578>
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19(2), 294–304. <https://doi.org/10.1037/0893-3200.19.2.294>
- Davis, J., Anderson, C., & Parker, W. (2019). Identifying and supporting Black male students in advanced mathematics courses throughout the K-12 pipeline. *Gifted Child Today*, 42(3), 140–149. <https://doi.org/10.1177/1076217519842234>
- Downey, D. B., von Hippel, P. T., & Hughes, M. (2008). Are “failing” schools really failing? Removing the influence of non-school factors from measures of school quality. *Sociology of Education*, 81(3), 242–270. <https://www.researchgate.net/publication/251422064>
- Drotos, S. M., Cilesiz, S., & English, S. (2016). Shoes, dues, and other barriers to college attainment: Perspectives of students attending high-poverty, urban high schools. *Education and Urban Society*, 48(3), 221–244. <https://doi.org/10.1177/0013124514533793>
- Duncan, G. J., & Murnane, R. J. (2014). Growing income inequality threatens American education. *Phi Delta Kappan*, 95(6), 8–14. <https://doi.org/10.1177/003172171409500603>
- Dusenbury, L., Dermody, C., & Weissberg, R. P. (2018). *2018 state scorecard scan*.
- Eccles, J. S., & Roeser, R. W. (1999). School and community influences on human development. In M. Bornstein & M. Lamb (Eds.), *Developmental Psychology: An Advanced Textbook* (4th ed., pp. 503–554). Lawrence Erlbaum.
- Eccles, J. S., & Roeser, R. W. (2008). Schools , academic motivation , and stage – environment fit. *Handbook of Adolescent Psychology*, 1, 404–434.
- Feng, L., Figlio, D., & Sass, T. (2018). School accountability and teacher mobility. *Journal of Urban Economics*, 103, 1–17. <https://doi.org/10.1016/j.jue.2017.11.001>
- Foster-Fishman, P. G., & Behrens, T. R. (2007). Systems change reborn: Rethinking our

- theories, methods, and efforts in human services reform and community-based change. *American Journal of Community Psychology*, 39(3–4), 191–196. <https://doi.org/10.1007/s10464-007-9104-5>
- Frisch, A. M. (2017). The class is greener on the other side: How private donations to public schools play into fair funding. *Duke Law Journal*, 67(2), 427–479. <https://doi.org/10.2139/ssrn.2915479>
- Gaddis, S. M., & Lauen, D. L. (2014). School accountability and the black–white test score gap. *Social Science Research*, 44, 15–31. <https://doi.org/10.1016/j.ssresearch.2013.10.008>
- Garnett, N. S. (2014). Disparate Impact, School Closures, and Parental Choice. *University of Chicago Legal Forum*, 2014, 289–344. https://scholarship.law.nd.edu/law_faculty_scholarship/1135
- Gershenson, S., Jacknowitz, A., & Brannegan, A. (2017). *Are student absences worth the worry in U.S. primary schools?* https://doi.org/10.1162/EDFP_a_00207
- Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, 31(4), 392–415. <https://doi.org/10.3102/0162373709342467>
- Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal*, 47(2), 434–465. <https://doi.org/10.3102/0002831209350494>
- Gouvea, J. S. (2021). Antiracism and the problems with “Achievement Gaps” in STEM education. *CBE—Life Sciences Education*, 20(1), fe2. <https://doi.org/10.1187/cbe.20-12-0291>
- Greenstone, M., Looney, A., Patashnik, J., & Yu, M. (2013). *Thirteen economic facts about social mobility and the role of education* (Issue June).
- Gregory, A., Cornell, D., & Fan, X. (2011). The relationship of school structure and support to suspension rates for Black and White high school students. *American Educational Research Journal*, 48(4), 904–934. <https://doi.org/10.3102/0002831211398531>
- Grissom, J. A., Rodriguez, L. A., & Kern, E. C. (2017). Teacher and principal diversity and the representation of students of color in gifted programs: Evidence from national data. *Elementary School Journal*, 117(3), 396–422. <https://doi.org/10.1086/690274>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Prentice-Hall, Inc.
- Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2012). *Achievement Growth: International and U.S. State Trends in Student Performance*.
- Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing

- achievement gap: Lessons from the past half-century. *American Journal of Education*, 112(2), 209–238. <https://doi.org/10.1086/498995>
- Harwell, M., Maeda, Y., Bishop, K., & Xie, A. (2017). The surprisingly modest relationship between SES and educational achievement. *The Journal of Experimental Education*, 85(2), 197–214. <https://doi.org/10.1080/00220973.2015.1123668>
- Hedges, L. V. (1981). Distribution theory for Glass’s estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.
- Hegedus, A. (2018). *Evaluating the Relationships Between Poverty and School Performance*.
- Hemelt, S. W., Ladd, H. F., & Clifton, C. R. (2021). Do teacher assistants improve student outcomes? Evidence from school funding cutbacks in North Carolina. *Educational Evaluation and Policy Analysis*, 43(2), 280–304. <https://doi.org/10.3102/0162373721990361>
- Hochschild, J. L., & Shen, F. X. (2014). Race, ethnicity, and education policy. In *Oxford Handbook of Racial and Ethnic Politics in America* (pp. 1–30). Oxford University Press. <https://doi.org/10.2139/ssrn.2476048>
- Hough, H., Kalogrides, D., & Loeb, S. (2017). *Using Surveys of Students’ Social-Emotional Learning and School Climate for Accountability and Continuous Improvement*.
- Huerta, A. H., Howard, T. C., & Haro, B. N. (2020). Supporting Black and Latino boys in school: A call to action. *Phi Delta Kappan*, 102(1), 29–33. <https://doi.org/10.1177/0031721720956846>
- Hughes, C., Bailey, C. M., Warren, P. Y., & Stewart, E. A. (2020). “Value in diversity”: School racial and ethnic composition, teacher diversity, and school punishment. *Social Science Research*, 92(August 2019), 102481. <https://doi.org/10.1016/j.ssresearch.2020.102481>
- Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., & Mary, J. C. (2020). The Condition of Education 2020 (NCES 2020-144). In *National Center for Educational Statistics at IES*. <https://nces.ed.gov/pubs2017/2017144.pdf>
- Ingersoll, R., Merrill, L., & May, H. (2016). Do Accountability Policies Push Teachers Out? *Educational Leadership*, 73(8), 44–49. https://repository.upenn.edu/gse_pubs/551
- Jacob, B. A. (2007). *Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments*. <http://www.nber.org/papers/w12817>
- Johnston, W. R., Engberg, J., Oppen, I. M., Sontag-Padilla, L., & Xenakis, L. (2020). *Illustrating the Promise of Community Schools: An Assessment of the Impact of the New York City Community Schools Initiative*.
- Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school

- accountability measures. *Journal of Economic Perspectives*, 16(4), 91–114.
<http://www.cde.ca.gov/psaa/api/>
- Kenny, M. E., Gualdron, L., Scanlon, D., Sparks, E., Blustein, D. L., & Jernigan, M. (2007). Urban adolescents' constructions of supports and barriers to educational and career attainment. *Journal of Counseling Psychology*, 54(3), 336–343.
<https://doi.org/10.1037/0022-0167.54.3.336>
- Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3–13.
<https://doi.org/10.3102/0013189X034008003>
- Kubota, Y., Heiss, G., Maclehose, R. F., Roetker, N. S., & Folsom, A. R. (2017). Association of educational attainment with lifetime risk of cardiovascular disease the atherosclerosis risk in communities study. *JAMA Internal Medicine*, 177(8), 1165–1172.
<https://doi.org/10.1001/jamainternmed.2017.1877>
- Lacour, M., & Tissington, L. D. (2011). The effects of poverty on academic achievement. *Educational Research and Reviews*, 6(7), 522–527. <http://www.academicjournals.org/ERR>
- Laditka, J. N., & Laditka, S. B. (2016). Associations of educational attainment with disability and life expectancy by race and gender in the United States: A longitudinal analysis of the Panel Study of Income Dynamics. *Journal of Aging and Health*, 28(8), 1403–1425.
<https://doi.org/10.1177/0898264315620590>
- Lane, K. L., Wehby, J. H., Robertson, E. J., & Rogers, L. A. (2007). How do different types of high school students respond to schoolwide positive behavior support programs? Characteristics and responsiveness of teacher-identified students. *Journal of Emotional and Behavioral Disorders*, 15(1), 3–20.
- Langford, I. H., & Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161(2), 121–160. <https://doi.org/10.1111/1467-985X.00094>
- Lapan, R. T., Gysbers, N. C., Stanley, B., & Pierce, M. E. (2012). Missouri Professional School Counselors: Ratios matter, especially in high-poverty schools. *Professional School Counseling*, 16(2), 108–116. <https://doi.org/10.1177/2156759X0001600207>
- Lawless, N. M., & Lucas, R. E. (2011). Predictors of regional well-being: A county level analysis. *Social Indicators Research*, 101(3), 341–357. <https://doi.org/10.1007/s11205-010-9667-7>
- Lee, J., & Lubienski, C. (2017). The impact of school closures on equity of access in Chicago. *Education and Urban Society*, 49(1), 53–80. <https://doi.org/10.1177/0013124516630601>
- Lenhoff, S. W., Pogodzinski, B., Mayrowetz, D., Superfine, B. M., & Umpstead, R. R. (2018). District stressors and teacher evaluation ratings. *Journal of Educational Administration*, 56(2). <https://doi.org/10.1108/JEA-06-2017-0065>

- Levernier, W., Partridge, M. D., & Rickman, D. S. (2000). The causes of regional variations in U.S. poverty: A cross-county analysis. *Journal of Regional Science*, 40(3), 473–497. <https://doi.org/10.1111/0022-4146.00184>
- Lewallen, T. C., Hunt, H., Potts-Datema, W., Zaza, S., & Giles, W. (2015). The Whole School, Whole Community, Whole Child Model: A new approach for improving educational attainment and healthy development for students. *Journal of School Health*, 85(11), 729–739. <https://doi.org/10.1111/josh.12310>
- Leyden, K. M. (2003). Social capital and the built environment: The importance of walkable neighborhoods. *American Journal of Public Health*, 93(9), 1546–1551. <https://doi.org/10.2105/AJPH.93.9.1546>
- Logan, S. R. (2018). A historical and political look at the modern school choice movement. *International Journal of Educational Reform*, 27(1), 2–21.
- Maier, A., Daniel, J., Oakes, J., & Lam, L. (2017). Community schools as an effective school improvement strategy: A Review of the evidence. *Learning Policy Institute*, December, 1–159. https://learningpolicyinstitute.org/product/%0Ahttps://learningpolicyinstitute.org/sites/default/files/product-files/Community_Schools_Effective_REPORT.pdf
- McGuinn, P. (2016). From No Child Left behind to the Every Student Succeeds Act: Federalism and the education legacy of the Obama Administration. *Publius: The Journal of Federalism*, 46(3), 392–415. <https://doi.org/10.1093/publius/pjw014>
- MDE. (2017). *Michigan's Consolidated State Plan Under the Every Student Succeeds Act*. <https://www.isbe.net/Documents/ESSAStatePlanforIllinois.pdf>
- MDE. (2018a). *2018-19 Public Guide to Michigan School Accountability Under the Every Student Succeeds Act (ESSA)*. https://www.michigan.gov/documents/mde/Public_Guide_to_Michigan_School_Accountability_618138_7.pdf
- MDE. (2018b). *Spring 2018 Interpretive Guide to M-STEP Reports*.
- MDE. (2019a). *Michigan School Index Results: Policy Considerations and Long-Term Educational Goals*. https://www.michigan.gov/documents/mde/MI_School_Index_Results_Policy_Considerations_Long_Term_Educ_Goals_660889_7.pdf
- MDE. (2019b). *Technical Report: Spring 2019 Michigan Student Test of Educational Progress (M-STEP)*.
- MDE. (2021). *Guide to State Assessments 2021-2022*. <https://www.michigan.gov/mde/Services/Student-Assessment>
- Michigan's Center for Educational Performance and Information (CEPI). (n.d.). Grades 3-8

- Assessments, High School Assessments, Attendance, Graduation/Dropout for All ISDs, All School Districts, All Schools, All Grades and All Students (2018-2019). <https://mischooldata.org/k-12-data-files/>. Retrieved March 10, 2022.
- Milner IV, H. R. (2013). Rethinking achievement gap talk in Urban Education. *Urban Education*, 48(1), 3–8. <https://doi.org/10.1177/0042085912470417>
- Mintrop, H., & Sunderman, G. L. (2009). Predictable failure of federal sanctions-driven accountability for school improvement-and why we may retain it anyway. *Educational Researcher*, 38(5), 353–364. <https://doi.org/10.3102/0013189X09339055>
- Mordechay, K., & Orfield, G. (2017). Demographic transformation in a policy vacuum: The changing face of U.S. metropolitan society and challenges for public schools. *The Educational Forum*, 81(2), 193–203. <https://doi.org/10.1080/00131725.2017.1280758>
- Noltemeyer, A. L., Ward, R. M., & McLoughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, 44(2), 224–240. <https://doi.org/10.17105/spr-14-0008.1>
- Owens, A., Reardon, S. F., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal*, 53(4), 1159–1197. <https://doi.org/10.3102/0002831216652722>
- Pascoe, J. M., Wood, D. L., Duffee, J. H., & Kuo, A. (2016). Mediators and adverse effects of child poverty in the United States. *Pediatrics*, 137(4). <https://doi.org/10.1542/peds.2016-0340>
- Patnode, A. H., Gibbons, K., & Edmunds, R. (2018). Attendance and chronic absenteeism: Literature review. *Center for Applied Research and Educational Improvement*, October, 51. www.cehd.umn.edu/CAREI/
- Pederson, P. V. (2007). What is measured is treasured: The impact of the No Child Left Behind Act on nonassessed subjects. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 80(6), 287–291. <https://doi.org/10.3200/TCHS.80.6.287-291>
- Peirson, L. J., Boydell, K. M., Ferguson, H. B., & Ferris, L. E. (2011). An ecological process model of systems change. *American Journal of Community Psychology*, 47(3–4), 307–321. <https://doi.org/10.1007/s10464-010-9405-y>
- Pena-Shaff, J. B., Bessette-Symons, B., Tate, M., & Fingerhut, J. (2019). Racial and ethnic differences in high school students' perceptions of school climate and disciplinary practices. *Race Ethnicity and Education*, 22(2), 269–284. <https://doi.org/10.1080/13613324.2018.1468747>
- Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137–1162. <https://doi.org/10.1177/016146811011200401>

- Pfeffer, F. T., & Hertel, F. R. (2015). How has educational expansion shaped social mobility trends in the United States? *Social Forces*, 94(1), 143–180.
<https://doi.org/10.1093/sf/sov045>
- Rappaport, J. (1987). Terms of empowerment/exemplars of prevention: Toward a theory for community psychology. *American Journal of Community Psychology*, 15(2), 121–148.
<https://doi.org/10.1007/BF00919275>
- Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances* (pp. 91–116). Russell Sage Foundation.
- Reardon, S. F. (2016). School district socioeconomic status, race, and academic achievement. In *Stanford Center for Education Policy Analysis*.
- Roda, A., & Wells, A. S. (2013). School choice policies and racial segregation: Where white parents' good intentions, anxiety, and privilege collide. *American Journal of Education*, 119(2), 261–293. <https://doi.org/10.1086/668753>
- Rowan, B., Barnes, C., & Camburn, E. (2004). Benefitting from comprehensive school reform: A review of research on CSR implementation. In C. Cross (Ed.), *Putting the pieces together: Lessons from comprehensive school reform research* (pp. 1–52). The National Clearinghouse for Comprehensive School Reform.
- Rumberger, R. W., & Losen, D. J. (2016). *The High Cost Of Harsh Discipline And Its Disparate Impact*. <https://escholarship.org/uc/item/85m2m6sj>
- Schneider, J., Noonan, J., White, R. S., Gagnon, D., & Carey, A. (2021). Adding “Student Voice” to the mix: Perception surveys and state accountability systems. *AERA Open*, 7(1), 233285842199072. <https://doi.org/10.1177/2332858421990729>
- Simpson, R. L., LaCava, P. G., & Sampson Graner, P. (2004). The No Child Left Behind Act: Challenges and implications for educators. *Intervention in School and Clinic*, 40(2), 67–75.
www.w-w-c.org
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453.
- Smart, M., Felton, J., Meghea, C., Buchalski, Z., Maschino, L., & Sadler, R. (2021). Is a school's neighborhood physical disorder related to its academic outcomes? *Child & Youth Care Forum*, 50(2), 247–259. <https://doi.org/10.1007/s10566-020-09572-3>
- Snellman, K., Silva, J. M., Frederick, C. B., & Putnam, R. D. (2015). The Engagement Gap. *The ANNALS of the American Academy of Political and Social Science*, 657(1), 194–207.
<https://doi.org/10.1177/0002716214548398>
- Sprouse, M. L. (2017). The consequential validity of the M-STEP and third-grade retention.

- Language Arts Journal of Michigan*, 32(2), 52–59. <https://doi.org/10.9707/2168-149X.2140>
- Strunk, K. O., Cowen, J. M., Torres, C., Burns, J., Waldron, S., & Auletto, A. (2019). *Partnership Turnaround : Year One Report* (Issue October).
- Sunderman, G. L., Coghlan, E., & Mintrop, R. (2017). *School closure as a strategy to remedy low performance*.
- The State School Aid Act of 1979a. MCL § 105. (1979).
[http://www.legislature.mi.gov/\(S\(rjbbvdoxdcyam13tkvu3nbtu\)\)/mileg.aspx?page=GetObject&objectname=mcl-388-1705](http://www.legislature.mi.gov/(S(rjbbvdoxdcyam13tkvu3nbtu))/mileg.aspx?page=GetObject&objectname=mcl-388-1705)
- The State School Aid Act of 1979b. MCL § 105c. (1979).
[http://www.legislature.mi.gov/\(S\(ljoq4dz4bmhyvpjochsmcyue\)\)/mileg.aspx?page=GetObject&objectname=mcl-388-1705c](http://www.legislature.mi.gov/(S(ljoq4dz4bmhyvpjochsmcyue))/mileg.aspx?page=GetObject&objectname=mcl-388-1705c)
- Tieken, M. C., & Auldridge-Reveles, T. R. (2019). Rethinking the school closure research: School closure as spatial injustice. *Review of Educational Research*, 89(6), 917–953.
<https://doi.org/10.3102/0034654319877151>
- Trickett, E. J. (1984). Toward a distinctive Community Psychology: An ecological metaphor for the conduct of community research and the nature of training. *American Journal of Community Psychology*, 12(3), 261–279.
- Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology*, 39(3–4), 217–228.
<https://doi.org/10.1007/s10464-007-9101-8>
- Turner, M., Kubatzky, L., & Jones, L. E. (2018). *Assessing Essa Missed Opportunities For Students With Disabilities*. https://hecse.net/wp-content/uploads/2020/09/18.-AssessingESSA_2018.pdf
- US Census Bureau (2023, February 9). *CPS Historical Time Series Tables*. *Census.gov*. Retrieved October 24, 2022, from <https://www.census.gov/data/tables/time-series/demo/educational-attainment/cps-historical-time-series.html>
- van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5(2), 134–150.
<https://doi.org/10.1016/j.edurev.2010.02.001>
- Waasdorp, T. E., Pas, E. T., O'brennan, L. M., & Bradshaw, C. P. (2011). A multilevel perspective on the climate of bullying: Discrepancies among students, school staff, and parents. *Journal of School Violence*, 10(2), 115–132.
<https://doi.org/10.1080/15388220.2010.539164>
- Warren, C. A., Douglas, T. R. M. O., & Howard, T. C. (2016). In their own words: Erasing deficits and exploring what works to improve K-12 and postsecondary Black male school achievement. *Teachers College Record*, 118(6), 1–6.

<https://doi.org/10.1177/016146811611800607>

- Winters, M. A. (2015). Understanding the gap in Special Education enrollments between charter and traditional public schools: Evidence from Denver, Colorado. *Educational Researcher*, 44(4), 228–236. <https://doi.org/10.3102/0013189X15584772>
- Xie, Y., Fang, M., & Shauman, K. (2015). STEM Education. *Annual Review of Sociology*, 41(1), 331–357. <https://doi.org/10.1146/annurev-soc-071312-145659>

APPENDIX A: LIST OF VARIABLES AND SOURCES

Table A1

List of Variables, Sources, and Years Available

Years	Generated Variables
Constant	Underperforming or Matched Comparison (district level)
Years	MI School Data Variables
2009-10 to 2018-19	Number of expulsions (school level)
2009-10 to 2018-19	Student counts (number enrolled; school level) broken down by: • Grade • Race/ethnicity • FRL status • IEP status
2009-10 to 2018-19	Student count mobile (school level) broken down by: • Race/ethnicity • FRL status • IEP status
2009-10 to 2018-19	Student count incoming (school level): • Race/ethnicity • FRL status • IEP status
2009-10 to 2018-19	Staffing counts and FTE (school level) broken down by: • Type of staff (Administrators, Paras, Subs, Teachers) • Race/ethnicity of staff • Gender of staff
2011-12 to 2018-19	Financial transparency (district level): • Expenditures • Revenue • Fund Balance • Years in Deficit • Resident students leaving • Non-resident students coming • Fund balance change
2011-12 to 2018-19	Attendance (school level): • Overall attendance rate by: ○ Grade (all grades) ○ Race/ethnicity ○ FRL status ○ IEP status
2014-15 to 2018-19	Assessment Data (school level) M-STEP ELA: • Number assessed • Counts & percentiles of proficiency

Table A1 (cont'd)

	• Mean scale scores (and SD) • Broken down by: ○ Grade (3-8) ○ Race/ethnicity ○ FRL status ○ IEP status M-STEP Math: • Number assessed • Counts & percentiles of proficiency • Mean scale scores (and SD) • Broken down by: ○ Grade (3-8) ○ Race/ethnicity ○ FRL status ○ IEP status
2015-16 to 2018-19	Student Growth (school level) ELA Growth Percentiles: • Number included • Number & percent above average, average, and below average growth • Mean SGP • Broken down by: ○ Grade (4-8) ○ Race/ethnicity ○ FRL status ○ IEP status Math Growth Percentiles: • Number included • Number & percent above average, average, and below average growth • Mean SGP • Broken down by: ○ Grade (4-8) ○ Race/ethnicity ○ FRL status ○ IEP status
Years	Civil Rights Data Collection (CRDC)
2009-10 to 2017-18 (every other year)	Number of counselors (school level)
2015-16 to 2017-18 (every other year)	Number of (school level): • Nurses • Psychologists • Social workers

Table A1 (cont'd)

2009-10 to 2017-18 (every other year)	Suspensions (school level) In-school suspensions broken down by: • Race/ethnicity • 504 status Out-of-school suspensions broken down by: • Race/ethnicity • 504 status
--	---

APPENDIX B: PERFORMANCE LEVELS FOR THE M-STEP

Table B1

Performance Levels for the 2018 Administration of the M-STEP

English Language Arts				
<u>Grade</u>	<u>Not Proficient</u>	<u>Partially Proficient</u>	<u>Proficient</u>	<u>Advanced</u>
3	1203-1279	1280-1299	1300-1316	1317-1357
4	1301-1382	1383-1399	1400-1416	1417-1454
5	1409-1480	1481-1499	1500-1523	1524-1560
6	1508-1577	1578-1599	1600-1623	1624-1655
7	1618-1678	1679-1699	1700-1725	1726-1753
8	1721-1776	1777-1799	1800-1827	1828-1857
Math				
<u>Grade</u>	<u>Not Proficient</u>	<u>Partially Proficient</u>	<u>Proficient</u>	<u>Advanced</u>
3	1217-1280	1281-1299	1300-1320	1321-1361
4	1310-1375	1376-1399	1400-1419	1420-1455
5	1409-1477	1478-1499	1500-1514	1515-1550
6	1518-1578	1579-1599	1600-1613	1614-1650
7	1621-1678	1679-1699	1700-1715	1716-1752
8	1725-1779	1780-1799	1800-1814	1815-1850

APPENDIX C: SUPPLEMENTAL TABLES

Table C1

School Total Enrollment by Partnership Status Over Time

School Year	Partnership District Schools		Matched Comparison District Schools	
	<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	461.62	330.34	525.53	395.72
2010-11	479.25	321.89	537.98	402.95
2011-12	468.34	306.10	528.36	403.09
2012-13	474.42	308.69	523.73	398.04
2013-14	484.12	311.86	502.95	399.68
2014-15	468.89	315.46	530.21	402.34
2015-16	455.30	312.66	507.55	409.13
2016-17	445.89	311.61	501.55	414.42
2017-18	445.55	313.78	499.99	418.26
2018-19	432.26	308.60	500.48	410.46

Table C2*School Demographics by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		Mean	SD	Mean	SD
2009-10	White Students	17.29%	22.90%	70.52%	23.03%
	Latine Students	8.75%	16.68%	7.41%	10.29%
	Black Students	67.66%	33.25%	17.90%	18.89%
	FRL Students	79.26%	16.09%	59.36%	19.87%
	IEP Students	20.82%	20.64%	13.95%	15.12%
2010-11	White Students	17.04%	21.84%	69.25%	23.00%
	Latine Students	12.47%	20.04%	7.94%	11.20%
	Black Students	66.59%	32.95%	17.88%	18.96%
	FRL Students	80.27%	13.85%	59.79%	20.14%
	IEP Students	19.05%	19.99%	13.83%	14.21%
2011-12	White Students	17.52%	21.50%	69.11%	22.95%
	Latine Students	12.95%	20.31%	8.08%	10.98%
	Black Students	65.48%	32.66%	17.10%	18.36%
	FRL Students	81.18%	12.87%	64.03%	18.93%
	IEP Students	20.35%	20.15%	14.22%	13.65%
2012-13	White Students	17.92%	21.85%	68.27%	23.36%
	Latine Students	13.66%	21.01%	8.52%	11.28%
	Black Students	64.16%	32.93%	16.90%	18.12%
	FRL Students	81.28%	13.17%	63.97%	18.18%
	IEP Students	18.81%	17.11%	13.73%	15.13%
2013-14	White Students	16.91%	20.62%	67.94%	23.28%
	Latine Students	13.66%	21.08%	8.83%	11.33%
	Black Students	65.04%	32.26%	16.27%	17.59%
	FRL Students	81.24%	12.89%	64.60%	17.83%
	IEP Students	17.47%	14.93%	13.62%	15.27%
2014-15	White Students	17.22%	21.31%	67.30%	23.62%
	Latine Students	14.02%	21.58%	9.26%	11.89%
	Black Students	64.10%	32.43%	16.06%	17.21%

Table C2 (cont'd)

	FRL Students	77.00%	14.89%	63.81%	18.67%
	IEP Students	17.26%	15.34%	12.68%	12.03%
2015-16	White Students	16.46%	20.40%	66.31%	24.37%
	Latine Students	14.15%	21.53%	9.31%	11.75%
	Black Students	64.53%	32.19%	16.55%	18.21%
	FRL Students	76.42%	14.49%	62.94%	17.29%
	IEP Students	17.63%	15.39%	13.40%	14.04%
2016-17	White Students	15.54%	19.24%	64.98%	24.54%
	Latine Students	14.18%	21.44%	10.04%	12.36%
	Black Students	65.48%	31.48%	16.36%	17.44%
	FRL Students	77.98%	14.12%	63.00%	18.40%
	IEP Students	17.78%	15.42%	12.86%	11.97%
2017-18	White Students	15.09%	18.73%	63.65%	25.27%
	Latine Students	14.84%	22.13%	10.12%	12.32%
	Black Students	64.86%	31.60%	16.80%	17.94%
	FRL Students	82.91%	14.02%	70.61%	17.50%
	IEP Students	17.99%	16.32%	13.80%	14.00%
2018-19	White Students	15.50%	19.13%	63.69%	24.85%
	Latine Students	15.38%	22.24%	10.26%	12.42%
	Black Students	63.62%	31.80%	16.36%	17.10%
	FRL Students	81.82%	14.34%	69.99%	16.89%
	IEP Students	17.82%	15.13%	13.45%	12.11%

Table C3*School Mobility Rate by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	All Students	36.56	21.40	20.36	16.03
	Black Students	37.69	23.01	30.35	21.76
	White Students	40.52	30.14	19.78	16.92
	Latine Students	38.26	34.91	21.96	21.76
	FRL Students	39.69	25.05	26.16	19.59
	IEP Students	38.40	16.04	34.32	19.30
2010-11	All Students	27.37	17.14	16.21	18.53
	Black Students	28.13	18.90	22.34	21.73
	White Students	27.44	26.76	15.26	19.30
	Latine Students	21.97	26.46	15.31	21.01
	FRL Students	29.80	18.97	18.22	18.64
	IEP Students	34.67	15.63	29.84	21.65
2011-12	All Students	16.62	15.67	10.35	13.03
	Black Students	17.71	16.10	14.98	16.63
	White Students	16.81	19.53	9.85	13.11
	Latine Students	13.04	19.39	12.32	16.65
	FRL Students	17.07	15.55	11.75	12.99
	IEP Students	22.02	14.58	18.90	13.48
2012-13	All Students	20.55	13.14	14.18	15.09
	Black Students	21.96	13.47	19.23	17.38
	White Students	25.67	22.93	13.62	15.98
	Latine Students	17.25	21.84	14.79	17.50
	FRL Students	21.55	13.19	16.43	15.25
	IEP Students	26.16	11.55	24.40	17.66
2013-14	All Students	14.39	10.77	9.01	10.52
	Black Students	15.94	12.82	11.43	15.26
	White Students	15.74	18.11	8.98	11.46

Table C3 (cont'd)

	Latine Students	13.64	18.71	9.21	13.25
	FRL Students	15.00	10.46	10.06	10.89
	IEP Students	19.60	11.82	18.09	15.40
2014-15	All Students	12.80	9.50	7.49	9.06
	Black Students	14.08	9.98	10.33	12.95
	White Students	17.53	20.12	7.16	9.13
	Latine Students	12.92	21.04	9.40	15.58
	FRL Students	13.44	9.48	8.86	10.13
	IEP Students	17.21	11.89	16.26	12.84
2015-16	All Students	12.88	9.92	9.09	12.62
	Black Students	13.73	10.62	12.16	15.56
	White Students	17.30	20.59	9.13	13.24
	Latine Students	12.12	17.96	10.47	17.50
	FRL Students	13.89	10.19	10.53	13.30
	IEP Students	18.00	6.66	19.14	21.78
2016-17	All Students	13.06	8.87	9.55	12.22
	Black Students	13.94	9.80	13.12	14.79
	White Students	17.91	20.59	9.06	12.33
	Latine Students	12.81	18.88	10.29	16.07
	FRL Students	13.85	8.90	10.94	12.78
	IEP Students	18.51	7.18	16.31	14.21
2017-18	All Students	13.93	9.98	10.01	13.99
	Black Students	14.97	10.57	13.32	19.04
	White Students	21.51	25.10	9.37	12.46
	Latine Students	12.23	19.57	10.31	17.75
	FRL Students	14.64	10.19	11.05	14.18
	IEP Students	19.18	8.01	20.03	20.09
2018-19	All Students	13.18	10.50	8.35	10.72
	Black Students	13.86	10.52	11.93	15.21
	White Students	17.55	20.18	8.10	11.06
	Latine Students	12.49	19.15	8.99	13.29

Table C3 (cont'd)

FRL Students	13.66	10.54	9.52	11.50
IEP Students	19.86	12.18	16.74	15.09

Table C4*District School of Choice by Partnership Status Over Time*

School Year	Category	Partnership Districts		Matched Comparison Districts	
		Mean	SD	Mean	SD
2011-12	% resident students leaving	.56	.24	.19	.12
	% non-resident students arriving	.10	.11	.09	.07
2012-13	% resident students leaving	.66	.35	.22	.12
	% non-resident students arriving	.13	.14	.10	.08
2013-14	% resident students leaving	.74	.41	.24	.13
	% non-resident students arriving	.13	.14	.11	.08
2014-15	% resident students leaving	.78	.44	.26	.14
	% non-resident students arriving	.13	.14	.12	.09
2015-16	% resident students leaving	.84	.51	.27	.16
	% non-resident students arriving	.15	.17	.12	.10
2016-17	% resident students leaving	.89	.56	.29	.17
	% non-resident students arriving	.16	.18	.13	.11
2017-18	% resident students leaving	.91	.57	.31	.18
	% non-resident students arriving	.16	.18	.15	.13
2018-19	% resident students leaving	.95	.61	.32	.20
	% non-resident students arriving	.18	.21	.14	.13

Table C5*School Staff per 100 Students by Partnership Status Over Time*

School Year	Category	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	Instructional paraeducators	2.47	4.00	1.65	2.69
	Non-instructional paraeducators	2.91	1.70	2.22	1.52
	Teachers	6.92	3.03	6.44	2.24
2010-11	Instructional paraeducators	2.49	4.20	1.75	3.02
	Non-instructional paraeducators	2.96	1.75	2.06	2.34
	Teachers	7.29	3.41	6.60	2.76
2011-12	Instructional paraeducators	2.50	4.00	1.69	3.26
	Non-instructional paraeducators	2.48	1.54	2.11	2.51
	Teachers	6.97	3.13	6.41	2.42
2012-13	Instructional paraeducators	2.75	4.48	2.20	7.51
	Non-instructional paraeducators	2.44	1.41	2.31	3.35
	Teachers	7.36	4.10	6.70	3.33
2013-14	Instructional paraeducators	2.46	3.95	2.44	9.36
	Non-instructional paraeducators	2.51	1.30	2.62	4.07
	Teachers	7.02	2.68	7.02	4.08
2014-15	Instructional paraeducators	2.44	3.84	1.79	4.12
	Non-instructional paraeducators	2.69	2.39	2.06	1.27
	Teachers	7.12	5.08	6.60	2.82
2015-16	Instructional paraeducators	2.53	3.91	1.75	3.70
	Non-instructional paraeducators	2.69	1.27	2.48	3.14
	Teachers	6.69	2.32	7.07	4.41
2016-17	Instructional paraeducators	2.49	3.52	1.73	3.49
	Non-instructional paraeducators	2.81	1.65	2.18	1.51
	Teachers	6.88	5.10	6.80	3.60
2017-18	Instructional paraeducators	2.32	3.32	1.89	3.68
	Non-instructional paraeducators	2.98	1.49	2.70	3.55
	Teachers	6.74	2.47	7.11	3.69

Table C5 (cont'd)

2018-19	Instructional paraeducators	2.34	3.48	1.94	3.99
	Non-instructional paraeducators	3.45	1.48	2.67	3.25
	Teachers	6.89	2.49	6.96	3.01

Table C6*Support Staff per 100 Students by Partnership Status Over Time*

School Year	Category	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	Counselors	.42	5.55	.08	.20
	Nurses	-	-	-	-
	Psychologists	-	-	-	-
	Social Workers	-	-	-	-
2011-12	Counselors	.29	.20	.32	.42
	Nurses	-	-	-	-
	Psychologists	-	-	-	-
	Social Workers	-	-	-	-
2013-14	Counselors	.15	.26	.14	.36
	Nurses	-	-	-	-
	Psychologists	-	-	-	-
	Social Workers	-	-	-	-
2015-16	Counselors	.15	.23	.12	.21
	Nurses	.08	.25	.04	.24
	Psychologists	.04	.09	.08	.21
	Social Workers	.10	.16	.17	.52
2017-18	Counselors	.19	.32	.17	.40
	Nurses	.06	.18	.04	.25
	Psychologists	.05	.10	.15	.60
	Social Workers	.09	.16	.27	.99

Table C7*Financial Characteristics (Dollars per Student) by Partnership Status Over Time*

School Year	Category	Partnership Districts		Matched Comparison Districts	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2011-12	Expenditures	\$13997.55	\$5299.33	\$9780.62	\$771.21
	Revenue	\$13755.43	\$5500.33	\$9801.75	\$746.76
	Fund Balance	\$-309.56	\$2392.19	\$812.64	\$756.46
2012-13	Expenditures	\$13391.36	\$5198.63	\$9676.85	\$782.34
	Revenue	\$13282.07	\$5864.50	\$9615.99	\$723.95
	Fund Balance	\$-748.52	\$3735.63	\$757.51	\$444.48
2013-14	Expenditures	\$13207.24	\$5929.16	\$9763.91	\$863.16
	Revenue	\$13300.56	\$5436.03	\$9739.25	\$730.45
	Fund Balance	\$-785.56	\$3288.00	\$732.79	\$559.37
2014-15	Expenditures	\$13533.38	\$6077.50	\$10095.44	\$858.25
	Revenue	\$13999.28	\$5957.05	\$10158.59	\$785.91
	Fund Balance	\$-365.93	\$3075.97	\$811.03	\$636.82
2015-16	Expenditures	\$14312.17	\$6672.82	\$10319.68	\$832.83
	Revenue	\$14748.32	\$6560.95	\$10464.60	\$751.41
	Fund Balance	\$-11.17	\$2870.45	\$969.84	\$663.73
2016-17	Expenditures	\$14407.60	\$7312.73	\$10554.36	\$775.77
	Revenue	\$15093.27	\$7148.88	\$10680.11	\$761.23
	Fund Balance	\$1137.89	\$2360.91	\$1121.63	\$699.50
2017-18	Expenditures	\$14646.73	\$6944.60	\$11118.23	\$858.97
	Revenue	\$15113.09	\$7016.78	\$11247.07	\$830.18
	Fund Balance	\$1632.11	\$1514.60	\$1282.83	\$756.55
2018-19	Expenditures	\$13364.46	\$2415.69	\$11259.82	\$819.45
	Revenue	\$13702.16	\$2295.93	\$11487.38	\$771.24
	Fund Balance	\$2008.77	\$1995.37	\$1510.66	\$900.11

Table C8*Attendance Rate by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2011-12	All Students	86.86	10.29	94.16	5.09
	Black Students	86.52	10.05	92.23	11.23
	White Students	81.01	24.63	94.42	4.93
	Latine Students	74.03	33.57	91.15	16.94
	FRL Students	86.67	10.16	93.67	5.13
	IEP Students	84.81	12.71	91.66	8.40
2012-13	All Students	88.63	8.60	93.76	5.15
	Black Students	88.20	8.54	92.08	8.97
	White Students	82.94	23.31	93.91	5.14
	Latine Students	79.47	29.89	91.79	13.96
	FRL Students	88.39	8.59	93.18	5.25
	IEP Students	86.32	13.20	91.48	11.37
2013-14	All Students	88.15	7.70	92.41	12.65
	Black Students	86.80	11.07	90.71	14.82
	White Students	83.99	21.57	92.61	12.69
	Latine Students	81.81	27.01	90.91	16.14
	FRL Students	87.84	7.67	91.78	12.75
	IEP Students	86.25	11.00	89.90	15.92
2014-15	All Students	88.18	7.09	93.29	7.34
	Black Students	87.71	6.95	91.66	7.63
	White Students	86.69	9.33	93.36	7.87
	Latine Students	88.48	9.14	92.28	8.76
	FRL Students	87.77	7.07	92.68	7.24
	IEP Students	86.68	7.52	91.76	6.92
2015-16	All Students	87.87	7.73	93.27	10.73
	Black Students	87.31	7.76	92.83	9.07
	White Students	89.17	7.66	93.40	10.74

Table C8 (cont'd)

	Latine Students	90.80	7.60	93.94	5.18
	FRL Students	87.40	7.86	92.85	10.48
	IEP Students	86.62	7.18	92.88	7.91
2016-17	All Students	88.10	6.98	93.12	9.53
	Black Students	87.63	6.69	92.43	8.77
	White Students	89.57	6.96	93.36	9.41
	Latine Students	91.58	6.32	92.62	9.83
	FRL Students	87.71	6.97	92.66	9.45
	IEP Students	86.73	6.22	92.96	7.41
2017-18	All Students	85.10	9.34	92.40	6.65
	Black Students	84.55	8.88	90.93	7.45
	White Students	87.78	9.51	92.58	7.00
	Latine Students	89.17	9.23	92.32	5.63
	FRL Students	84.69	9.28	91.82	6.63
	IEP Students	83.50	8.93	90.98	7.35
2018-19	All Students	86.42	8.59	91.70	8.28
	Black Students	85.84	8.44	90.04	8.21
	White Students	87.91	8.53	92.13	8.07
	Latine Students	89.47	8.75	91.76	7.73
	FRL Students	86.01	8.58	91.00	8.61
	IEP Students	85.38	7.16	90.51	6.86

Table C9*ELA Z Scores by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2014-15	All Students	-.25	.84	.94	.63
	Black Students	-.48	.71	.18	.49
	White Students	.40	.94	1.20	.62
	Latine Students	.03	.71	.47	.46
	FRL Students	-.32	.78	.62	.59
	IEP Students	-1.28	.52	-.58	.59
2015-16	All Students	-.27	.85	.79	.67
	Black Students	-.49	.69	.01	.53
	White Students	.38	1.02	1.04	.66
	Latine Students	-.02	.64	.40	.56
	FRL Students	-.34	.79	.49	.66
	IEP Students	-1.37	.44	-.60	.52
2016-17	All Students	-.31	.81	.70	.65
	Black Students	-.54	.68	-.06	.49
	White Students	.33	1.01	.97	.70
	Latine Students	.02	.67	.43	.43
	FRL Students	-.39	.76	.40	.59
	IEP Students	-1.33	.48	-.62	.62
2017-18	All Students	-.41	.78	.61	.69
	Black Students	-.67	.61	-.25	.53
	White Students	.29	1.05	.91	.66
	Latine Students	-.14	.62	.32	.45
	FRL Students	-.47	.72	.34	.64
	IEP Students	-1.40	.42	-.65	.45
2018-19	All Students	-.39	.78	.67	.72
	Black Students	-.64	.66	-.20	.59
	White Students	.30	1.01	.99	.68

Table C9 (cont'd)

Latine Students	-.09	.59	.29	.42
FRL Students	-.46	.72	.41	.69
IEP Students	-1.33	.46	-.60	.53

Table C10*Math Z Scores by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2014-15	All Students	-.29	.78	.84	.65
	Black Students	-.56	.65	-.04	.43
	FRL Students	-.36	.71	.54	.61
	IEP Students	-1.22	.50	-.50	.59
	Latine Students	-.01	.64	.47	.52
	White Students	.41	.84	1.11	.60
2015-16	All Students	-.35	.78	.75	.71
	Black Students	-.60	.63	-.11	.52
	FRL Students	-.41	.74	.47	.69
	IEP Students	-1.35	.52	-.56	.59
	Latine Students	-.03	.62	.40	.56
	White Students	.31	.90	1.03	.63
2016-17	All Students	-.33	.76	.77	.68
	Black Students	-.58	.66	-.13	.53
	FRL Students	-.41	.69	.48	.65
	IEP Students	-1.25	.49	-.50	.60
	Latine Students	.05	.62	.46	.48
	White Students	.33	.96	1.05	.65
2017-18	All Students	-.43	.78	.69	.70
	Black Students	-.72	.61	-.24	.52
	FRL Students	-.49	.72	.43	.67
	IEP Students	-1.52	.47	-.68	.59
	Latine Students	-.09	.63	.46	.50
	White Students	.30	1.02	1.03	.66
2018-19	All Students	-.38	.80	.75	.73
	Black Students	-.67	.66	-.20	.48
	FRL Students	-.44	.76	.51	.70

Table C10 (cont'd)

IEP Students	-1.44	.54	-.63	.60
Latine Students	.02	.60	.37	.48
White Students	.31	1.00	1.08	.66

Table C11*ELA Growth Percentiles by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2015-16	All Students	41.51	9.03	47.39	8.83
	Black Students	39.84	8.30	43.93	9.39
	White Students	46.15	10.51	48.42	8.71
	Latine Students	45.99	8.24	47.14	10.28
	FRL Students	41.06	8.97	46.38	8.71
	IEP Students	38.28	10.14	42.91	10.02
2016-17	All Students	44.27	8.76	48.76	8.11
	Black Students	43.11	8.99	46.26	7.93
	White Students	47.18	9.35	50.34	8.16
	Latine Students	48.55	9.21	48.30	9.01
	FRL Students	43.87	8.76	47.48	8.46
	IEP Students	40.78	8.63	43.91	9.55
2017-18	All Students	44.27	8.16	48.54	9.83
	Black Students	43.34	8.36	44.20	9.84
	White Students	47.09	9.50	50.18	8.90
	Latine Students	48.05	7.53	46.92	9.07
	FRL Students	43.96	8.07	47.59	9.71
	IEP Students	41.93	8.47	46.44	9.57
2018-19	All Students	44.29	7.10	47.90	9.61
	Black Students	42.93	6.98	44.43	9.76
	White Students	47.64	8.11	49.31	9.24
	Latine Students	47.81	7.73	47.11	9.34
	FRL Students	44.08	7.12	47.49	9.25
	IEP Students	39.55	8.82	43.51	9.31

Table C12*Math Growth Percentiles by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2015-16	All Students	40.41	7.95	47.22	9.41
	Black Students	39.06	7.02	44.46	9.09
	White Students	43.94	9.77	48.41	9.44
	Latine Students	44.08	8.25	47.32	10.58
	FRL Students	40.14	7.93	46.53	9.21
	IEP Students	39.15	9.28	43.94	10.53
2016-17	All Students	42.87	9.10	48.66	9.15
	Black Students	42.04	9.35	46.21	9.12
	White Students	45.73	10.32	50.23	9.00
	Latine Students	47.30	10.20	48.21	9.78
	FRL Students	42.59	9.04	47.86	8.61
	IEP Students	41.47	9.80	47.01	11.47
2017-18	All Students	42.37	8.68	47.79	9.67
	Black Students	41.64	8.18	44.74	8.84
	White Students	45.22	10.63	49.07	9.95
	Latine Students	45.68	8.85	48.15	10.10
	FRL Students	42.21	8.33	47.02	8.99
	IEP Students	39.15	9.51	44.74	9.36
2018-19	All Students	43.70	7.96	48.37	9.74
	Black Students	42.46	7.92	44.25	9.98
	White Students	45.75	9.91	50.01	9.35
	Latine Students	46.99	8.46	45.89	10.19
	FRL Students	43.45	7.93	47.79	10.12
	IEP Students	41.20	8.87	46.29	9.70

Table C13*Expulsions by Partnership Status Over Time*

School Year	Partnership District Schools		Matched Comparison District Schools	
	<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	.33	1.49	.74	2.29
2010-11	.44	1.35	.83	2.31
2011-12	.54	1.79	1.04	2.72
2012-13	.76	2.53	1.13	3.42
2013-14	.49	1.48	1.18	3.10
2014-15	.41	1.20	.83	2.48
2015-16	.21	1.03	.79	2.15
2016-17	.45	1.31	.93	3.29
2017-18	.35	.93	.62	1.83
2018-19	.62	1.68	.70	2.06

Table C14*In-School Suspensions by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	All Students	1.16	6.18	4.18	10.54
	IEP Students	.97	5.05	5.76	14.31
	White Students	.79	4.84	3.28	9.15
	Black Students	1.21	6.40	4.93	11.86
	Latine Students	.52	4.42	2.45	8.12
2011-12	All Students	1.14	4.18	5.20	11.02
	IEP Students	1.47	5.13	6.20	12.30
	White Students	1.06	4.03	4.20	9.91
	Black Students	1.44	5.59	8.46	15.63
	Latine Students	.59	3.03	4.87	10.28
2013-14	All Students	2.03	5.57	2.32	4.30
	IEP Students	3.08	8.08	4.12	8.28
	White Students	1.57	4.64	1.85	3.73
	Black Students	2.72	7.27	5.13	11.23
	Latine Students	1.31	4.60	3.74	9.19
2015-16	All Students	2.64	7.04	3.47	9.19
	IEP Students	3.19	8.79	3.70	10.12
	White Students	2.59	8.13	2.56	5.98
	Black Students	3.35	8.59	5.99	13.85
	Latine Students	1.72	5.22	2.62	7.26
2017-18	All Students	1.76	3.96	3.64	7.36
	IEP Students	2.41	5.51	5.54	12.13
	White Students	1.40	3.28	3.37	6.86
	Black Students	3.14	6.88	6.59	13.62
	Latine Students	1.08	3.56	3.01	7.50

Table C15*Out-of-School Suspensions by Partnership Status Over Time*

School Year	Student Subgroup	Partnership District Schools		Matched Comparison District Schools	
		<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
2009-10	All Students	17.19	19.30	8.84	11.96
	IEP Students	18.88	22.46	12.09	18.33
	White Students	5.38	13.20	6.56	10.17
	Black Students	19.58	21.47	12.12	16.99
	Latine Students	5.76	14.35	4.23	10.27
2011-12	All Students	19.95	16.48	9.23	12.03
	IEP Students	25.64	21.30	12.72	14.52
	White Students	16.14	19.65	7.57	10.81
	Black Students	22.84	18.21	16.49	18.21
	Latine Students	10.66	14.86	9.19	13.91
2013-14	All Students	16.77	12.31	7.49	8.37
	IEP Students	30.78	21.83	16.16	14.91
	White Students	15.16	18.16	7.68	10.08
	Black Students	25.78	18.24	17.81	19.52
	Latine Students	13.46	16.66	11.98	19.13
2015-16	All Students	20.81	15.28	9.98	13.02
	IEP Students	26.84	18.80	10.48	14.11
	White Students	16.15	18.17	7.95	10.25
	Black Students	24.67	16.62	16.08	16.33
	Latine Students	14.38	17.55	7.16	10.92
2017-18	All Students	20.04	16.03	10.73	13.05
	IEP Students	24.62	18.62	12.62	15.14
	White Students	16.25	15.24	8.82	11.46
	Black Students	26.22	18.47	16.54	17.47
	Latine Students	12.59	13.03	7.67	13.02