

education policy analysis archives

A peer-reviewed, independent,
open access, multilingual journal



Arizona State University

Volume 28 Number 89

June 8, 2020

ISSN 1068-2341

The Multidimensionality of School Performance: Using Multiple Measures for School Accountability and Improvement

*Min Sun*¹

University of Washington
United States

Alec I. Kennedy

San Francisco Unified School District
United States

&

Eric M. Anderson

Seattle Public Schools
United States

Citation: Sun, M., Kennedy, A. I., & Anderson, E. M. (2020). The multidimensionality of school performance: Using multiple measures for school accountability and improvement. *Education Policy Analysis Archives*, 28(89). <https://doi.org/10.14507/epaa.28.4689>

Abstract: The Every Student Succeeds Act of 2015 grants states and districts the flexibility to use multiple measures to assess school performance and strategically manage public schools for improvement in the United States. However, there is a lack of systematic,

¹ Min Sun's efforts on this study is supported by a grant from the National Science Foundation under Grant No. DRL-1506494. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

evidence-based guidance for practitioners on how to interpret the complex relationships between these multiple measures. Drawing on the organizational management literature on the multidimensionality of organizational effectiveness, along with longitudinal data from Washington State, we illustrate the multidimensionality of school performance and different measurement properties of school performance data. We also find that schools that are higher performing in terms of students' average scale scores and average growth percentiles in some cases have larger disparities in these same measures between historically underserved students of color and their peers than lower performing schools do. Moreover, these performance measures have time-series properties. The complexity of school performance measurement systems calls for continuous support for local educators to appropriately use school performance data to promote student success.

Keywords: school performance; school accountability; school improvement; multidimensionality; multiple measures

Las múltiples dimensiones del rendimiento escolar: Uso de múltiples medidas para la rendición de cuentas y la mejora de las escuelas

Resumen: The Every Student Succeeds Act of 2015 permite estados y distritos de los Estados Unidos la flexibilidad de utilizar múltiples medidas para evaluar el rendimiento escolar y administrar estratégicamente las escuelas públicas para mejorar. Sin embargo, existe una falta de orientación sistemática y basada en la evidencia para los profesionales sobre cómo interpretar las complejas relaciones entre estas medidas múltiples. Basándonos en la literatura sobre efectividad organizacional, junto con datos longitudinales del estado de Washington, ilustramos las múltiples dimensiones del rendimiento escolar y las diferentes propiedades de medición de los datos sobre el rendimiento escolar.

Encontramos que las escuelas con un rendimiento más alto en términos de puntajes de escala promedio de los estudiantes y percentiles de crecimiento promedio en algunos casos tienen mayores disparidades en estas mismas medidas entre los estudiantes de color y sus compañeros que las escuelas con un rendimiento más bajo. Estas medidas de rendimiento también tienen propiedades de series temporales. La complejidad de los sistemas que miden el rendimiento escolar requiere un apoyo continuo para que los educadores locales utilicen adecuadamente los datos para promover el éxito de los estudiantes.

Palabras-clave: rendimiento escolar; rendición de cuentas escolar; mejora escolar; multidimensionalidad; medidas múltiples

As múltiplas dimensões do desempenho escolar: Uso de várias medidas para a prestação de contas e melhoria das escolas

Resumo: The Every Student Succeeds Act of 2015 permite aos estados e distritos dos Estados Unidos a flexibilidade de usar várias medidas para avaliar o desempenho da escola e gerenciar estrategicamente as escolas públicas para melhoria. No entanto, há uma falta de orientação sistemática e baseada em evidências para os profissionais sobre como interpretar as complexas relações entre essas múltiplas medidas. Com base na literatura sobre eficácia organizacional, juntamente com dados longitudinais do estado de Washington, ilustramos as múltiplas dimensões do desempenho escolar e as diferentes propriedades de medição dos dados sobre o desempenho escolar. Concluímos que as escolas com melhor desempenho em termos de pontuação na escala média dos alunos e percentis de crescimento médio em alguns casos apresentam disparidades maiores nessas mesmas medidas entre estudantes de cor e seus colegas do que as escolas com

desempenho inferior. Essas medidas de desempenho também têm propriedades de séries temporais. A complexidade dos sistemas que medem o desempenho da escola exige suporte contínuo para que os educadores locais usem os dados adequadamente para promover o sucesso do aluno.

Palavras-chave: desempenho escolar; prestação de contas escolar; melhoria escolar; multidimensionalidade; medidas múltiplas

Introduction

The school performance measurement system established by the Every Student Succeeds Act (ESSA) of 2015 broadened the definition of school performance in the United States. Under the legislation that preceded it, No Child Left Behind (NCLB), school performance for accountability purposes was primarily assessed in terms of output-oriented measures, including student achievement, academic growth, and graduation rates. Under ESSA, in addition to these output-oriented measures, state school accountability systems can include input- and process-related measures associated with school efficiency or effectiveness, such as the availability of advanced coursework, teacher quality and its equitable distribution, teacher collaboration and school leadership, and student engagement (attendance) and behavior (discipline). Including such measures addresses a problem often discussed in the organizational management literature—that the performance measures used to evaluate the success of nonprofit organizations are often not linked to strategies for improvement (Sawhill & Williamson, 2001). With ESSA's introduction of a broadened measurement system for school performance, the intention was to provide districts and schools with more information that would help them identify strategies for improvement (Bae, 2018).

However, research has provided states and school districts with little guidance on how to incorporate multiple measures into one accountability system and how to use them for school improvement. One basic challenge is understanding the relationships between performance measures and what they mean for schools. Individual measures capture different aspects of school performance, and combining them into an overall summative score may mask which schools are lowest performing on certain measures and how schools should target those areas with their improvement efforts (Hough, Penner, & Witte, 2016). Furthermore, process-related measures—such as student attendance, student disciplinary infractions, and teacher collaboration and school leadership—on one hand may provide useful information for districts and schools as they make improvement plans but on the other hand may be less standardized, reliable, or valid for conducting cross-school or cross-district comparisons (Duckworth & Yeager, 2015; West, 2016). ESSA grants states and districts the flexibility to use different measures to develop local theories of change for school improvement, but to fully leverage this flexibility, policymakers and practitioners need to understand the complex relationships between these measures and the tradeoffs of using each.

Another challenge is understanding the relationship between schools' overall performance and their differential contributions to subgroups of students. Consistent with NCLB, ESSA legislation requires states to report measures disaggregated by student subgroup, including by race/ethnicity, income, English proficiency level, and disability status. Prior studies have highlighted disparities in learning opportunities and outcomes for racially underserved students of color or low-income students and provided basic methods for gauging educational inequality (e.g., Reardon, 2019; Reardon & Ho, 2015; Reardon & Robinson, 2008; Skiba et al., 2014). However, less is known about the relationship between a school's average performance and its students' differential performance—that is, within-school disparities in student achievement and opportunities to learn—in an accountability context.

To shed light on the complexity of using multiple measures for school accountability and improvement, in the current study, we use data from Washington State and its largest school district, Seattle Public Schools, to explore the following research questions:

1. What are the relationships among several student behavioral and academic measures of school performance, including schools' achievement levels, academic growth, graduation rates, chronic absenteeism rates, and disciplinary infraction rates? Some combination of these measures has been used in virtually every approved ESSA state accountability plan, and by examining the relationships between them, we are able to illustrate why and how states and school districts need to understand these measures and use them appropriately for accountability and continuous improvement.
2. How are disparities in student behavioral and academic measures for racial/ethnic subgroups related to schools' average performance? Do historically underserved students of color benefit as much as their peers do from schools that are high-performing on average?
3. How do schools' internal process measures (such as student and staff perceptions of school climate) explain variations in school performance and equity? Perceptions of school climate have been widely used by states and districts to measure school quality (e.g., in Alaska, Vermont, Chicago Public Schools, and New York City) or to inform school improvement (e.g., in Montana and the CORE school districts in California).

By examining these questions, our study makes several contributions to the literature on school accountability. First, although prior studies have illustrated several potential issues related to implementing multiple measures of school performance (Hough et al., 2016), researchers have not encapsulated these issues in a coherent framework. This study is the first, to our knowledge, to draw on the framework of multidimensional organizational effectiveness from the management literature to inform the understanding and measurement of school performance.

Second, although prior literature has foreshadowed several dimensions of the complexity of the school measurement system under ESSA, it has not empirically illustrated them in a school accountability and improvement system. For example, discussions of using multiple measures of teacher effectiveness in teacher evaluations have some bearing on measuring schools using multiple measures (Kane, McCaffrey, Miller, & Staiger, 2013). Another example can be found in methodological discussions of school value-added and student growth percentiles (e.g., Ehlert, Koedel, Parsons, & Podgursky, 2014; Raudenbush, 2009; Reardon & Raudenbush, 2009). Another example includes studies on disparities in student achievement and disciplinary infractions between racial/ethnic groups that provide basic knowledge and methods for gauging educational inequality across subgroups of students (e.g., Reardon & Ho, 2015; Skiba et al., 2014). However, none of these studies was conducted in the context of informing high-stakes school accountability measures such as those used under ESSA. The current study is unique in its use of real data from one state and one district to illustrate issues similar to those reflected in prior studies and synthesize recommendations into a coherent framework that practitioners can use to inform continuous improvement in schools.

Third, this paper is cowritten by researchers and a practitioner and adds a unique, reflective lens to the issues around implementing ESSA's school performance measurement system at the state and district levels. As described in a special issue of *Education Policy Analysis Archives*, ESSA makes state and local leaders fundamentally rethink how they approach school performance and accountability (Stosich, Bae, & Snyder, 2018). Our discussions of the premises and pitfalls of

multiple measures accountability systems provide actionable knowledge that can inform states' and districts' use of evidence for school improvement under ESSA.

The Multidimensionality of School Effectiveness and Its Measurement

To understand the ESSA school accountability measurement system from a conceptual standpoint, we draw on the organizational management literature on the multidimensionality of organizational effectiveness (Bentes, Carneiro, da Silva, & Kimura, 2012; Cameron, 1986; Chakravarthy, 1986; Herman & Renz, 2008; Richard, Devinney, Yip, & Johnson, 2009; Venkatraman & Ramanujam, 1986). This literature emphasizes that organizations—both profit and nonprofit—are inherently complex and should not be measured by a unidimensional metric. Strategic performance management requires organizations to use diverse measurement systems to track and manage local initiatives, connect these initiatives with organizational missions and goals, and efficiently meet the needs of their stakeholders.

The multidimensionality of organizational effectiveness means that different theoretical and empirical components of organizational performance may or may not be related, and yet they collectively reflect the health of an organization (Callen, 1991; Devinney, Yip, & Johnson, 2010; Henri, 2004). To inform organizational improvement, then, the measurement of organizational performance needs multiple discipline-specific measures that address the relationships between organizational inputs, processes, and outputs. The use of multiple measures is critical because research shows that firms with more extensive measurement systems—especially those that include objective financial and subjective stakeholder perception measures—perform better in terms of stock market returns (Ittner, Larcker, & Randall, 2003; Van der Stede, Chow, & Lin, 2006). The literature continues to suggest that to preserve the underlying dimensionality of performance measures, it is more effective to operate at the disaggregated level and not to impose relationships between measures unless there is a sound theoretical model for how the measures are related to one another.

Before illustrating measurement issues in operation, it is critical to understand the sources of the multidimensionality of organizational effectiveness. Richard and his colleagues (2009) conceptualized three sources of multidimensionality, all of which need to be accounted for to validly measure performance: (a) Who are the *stakeholders* for whom a performance measure is relevant? (b) What is the *landscape* over which performance is being determined? Organizations are heterogeneous in their resources, capabilities, and how and where they choose to use them, and measurements of performance must account for this heterogeneity; and (c) What *timeframe* is relevant in measuring performance? Drawing on this framework and relevant studies in education, then, we discuss four aspects of the multidimensionality of school performance.

First, *the stakeholders of public schools have diverse values and interests*. Scholars have elaborated the diverse and often conflicting values and ideals among stakeholders (Green, 1983). For example, state policymakers represent interests at the aggregate level in society. They may value promoting social and economic efficiency and equity as important outcomes of schooling. In contrast, individual parents may care more about their own child's academic performance and behavioral development and how schools can give their child an advantage in future employment and life (Green, 1983). Even among parents, some may place a higher value on academic performance, while others may care more about their children's behavioral development. Moreover, community stakeholders may expect schools to serve local needs, boosting economic development and solving societal problems. Recognizing students' need to master skills beyond basic competencies in tested subjects (e.g., math and English language arts), such as creativity, problem-solving, and the ability to manage their

behaviors and emotions, states and districts broaden their educational priorities and the associated metrics (Bae, 2018). For schools that aim to serve diverse interests and goals, it is only logical to assume that instruction and school practices serve multiple ends.

Second, *school performance varies depending on school resources, strategic choices, and policy environments*. The fiscal resources available to schools affect the services and reform strategies they adopt (Dragoset et al., 2017). The quantity and quality of intellectual resources that are accessible—such as instructional and leadership coaches who can serve as critical sources of knowledge and information—may also affect school practices. Schools' strategic choices subsequently affect school performance (Rumberger & Palardy, 2005; Sun, Saultz, & Ye, 2017). For example, schools that adopt positive behavioral intervention support for at-risk students may expect to see reductions in disciplinary infractions and absenteeism (e.g., Bohanon et al., 2006; Flannery, Fenning, Kato, & McIntosh, 2014; Freeman et al., 2015; Horner et al., 2009; Sun, Liu, Zhu, & LeClair, 2019). Other schools that adopt ambitious instructional approaches or expand the advanced course options available to their students (such as International Baccalaureate or Advanced Placement courses) may expect to see an uptick in student test scores, graduation rates, and college readiness (e.g., Burris, Heubert, & Levin, 2006; Kolluri, 2018). Moreover, the political context in which a school is situated also matters. Schools that are under federal or state accountability pressure to improve student achievement may focus their resources strategically on preparing students for standardized tests (Darling-Hammond & Adamson, 2014; Hamilton, Berends, & Stecher, 2005; Lauen & Gaddis, 2016; Stecher, Barron, Kaganoff, & Goodwin, 1998). Other schools that are audited for racial/ethnic disparities in student outcomes may regard educational equity as a central goal of their reform efforts and take strategic actions to reduce achievement gaps (Sparks, 2015). Therefore, to account for variations in school resources, strategic choices, and organizational contexts, a diverse set of measures is needed to provide actionable, local, and useful information to support school improvement.

Third, *schools may have heterogeneous contributions to different groups of students*. Prior literature in education provides ample evidence on disparities in learning opportunities and educational attainment among students from different racial/ethnic subgroups. Historically underserved students of color (including American Indian/Alaska Native, Black/African American, Hispanic/Latino, and Pacific Islander students) tend to have lower achievement and higher disciplinary infraction rates. Students from these groups are more likely to enroll in special education and less likely to enroll in gifted programs and Advanced Placement courses (e.g., Hosp & Reschly, 2004; Reardon, Kalogrides, & Shores, 2016; Reardon & Robinson, 2008; Shores, Kim, & Still, 2017; Wallace, Goodkind, Wallace, & Bachman, 2008). What is more, several school components that are critical to student learning may not be distributed equally. Even within schools, students who are members of historically underserved minority groups, are eligible for subsidized lunch, or have lower prior achievement are more likely to have teachers with fewer years of experience, lower licensure exam scores, and lower value-added scores (Goldhaber, Lavery, & Theobald, 2015; Kalogrides & Loeb, 2013; Kalogrides, Loeb, & Bêteille, 2013). Further, school practices such as tracking, patterned participation in extracurricular activities, and segregated peer networks can generate unequal learning opportunities for different groups of students (Carter, 2012; Crosnoe, 2009; Hallinan & Williams, 1989; Mickelson, 2003; Moody, 2002). Such differences in access to learning opportunities may explain why achievement gaps within schools account for more of the overall disparities in achievement than do gaps between schools (Bohrnstedt, Kitmitto, Ogut, Sherman, & Chan, 2015). Systems for measuring school performance need to attend to the possibility that schools may have differential contributions to students within schools.

Fourth, *the influence of school practices on student learning can be incremental*. Some measures used in states' ESSA accountability systems pertain to student outcomes (e.g., achievement), and others pertain to the processes or conditions for learning (e.g., attendance, positive learning environment in the classroom, effective teachers). The latter are often considered "leading indicators" of change, occurring before substantial changes in student achievement can be observed. Policymakers and researchers have leveraged this temporal relationship between measures to diagnose issues within schools and develop interventions. For instance, Sun et al.'s prior study in the San Francisco Unified School District found that reform efforts aimed at transforming underperforming schools showed early improvements in many process measures, including a reduction in unexcused student absences and an increase in the retention of high value-added teachers in math and English language arts (Sun et al., 2017). These improvements in process measures provide evidence on the gradual improvements in distal outcomes (such as test scores). However, it should be noted that short- or medium-term success may not always translate to long-run changes in outcomes. These short-term successes can be heavily biased by random fluctuations. Accordingly, researchers and school leaders should interrogate any process measures used by examining the degree to which they predict long-term outcomes.

In sum, as with organizational performance in other industries, the multidimensionality of school performance is rooted in several factors, including the diversity of stakeholder interests, strategic choices regarding initiatives and resource allocation, interactions between school services and student backgrounds, and the temporal relationship between organizational activities and outcomes. This multidimensionality requires that a diverse set of measures be used in assessing school performance. The ESSA measurement system intends to meet this requirement, yet our knowledge of the complex process of integrating multiple measures into one coherent system has been insufficient to fully guide practitioners in the design and implementation of school accountability plans. The current study helps address this gap in knowledge by using empirical data to illustrate the mutual consistency and predictive relationships between several measures of school performance that are typically included in states' ESSA plans.

Method

Data and Sample

To examine the relationships between several measures of school effectiveness and the disparities between racial/ethnic student groups, we use data from all public K-12 schools in the state of Washington from 2010 (the spring of the 2009–10 school year) to 2016. Since Washington State did not collect statewide school and teacher surveys longitudinally (and most states in the country do not), in our exploration of the role of malleable school processes in student outcomes, we use student and staff survey data from Seattle Public Schools from 2014 to 2016.

Our sample only includes one state and one district, so the findings may be limited to the context of this state or this district. Moreover, the measures we developed do not completely align with Washington State's school performance measures under ESSA, nor do they reflect exactly how Seattle Public Schools uses its data for improvement. Because our intent with this study is not to discuss one state's policy or one district's data use, in determining which measures to focus on, we draw inspiration from multiple states' school accountability designs and districts' possible ways of using data. Our aim is to identify patterns that inform state policymakers, district administrators, and researchers about issues to consider in measuring school performance for accountability and improvement.

Table 1
Descriptive Statistics

Characteristics	Washington State		Seattle Public Schools	
	Elementary/ Middle	High	Elementary/ Middle	High
Mean enrollment	483.36 (222.23)	548.37 (590.16)	471.84 (211.74)	691.28 (535.30)
% HUSC	28 (22)	28 (22)	32 (20)	41 (22)
% free/reduced-price lunch	51 (25)	50 (23)	41 (28)	47 (23)
% primary language is English	81 (20)	86 (18)	78 (20)	74 (23)
% homeless	3 (3)	4 (6)	3 (3)	7 (10)
% female	48 (4)	46 (10)	48 (3)	47 (8)
<i>N</i>	1,847	529	90	17

Note: HUSC = historically underserved students of color. Standard deviations are shown in parentheses.

As shown in Table 1, Washington State's public schools serve nearly 1.1 million students annually in more than 1,800 elementary and middle schools and 500 high schools. From 2010 to 2016, about 28% of students in the state were identified as historically underserved students of color (i.e., Hispanic/Latino, Black/African American, American Indian/Alaska Native, or Pacific Islander). Additionally, about half of students in the state were eligible for free or reduced-price lunch, almost 20% reported that their primary language was not English, and about 3% were identified as homeless.

Seattle Public Schools, the largest school district in Washington, enrolls almost 55,000 students and is similarly diverse. Between 2010 and 2016, 32% of Seattle Public School students were identified as historically underserved students of color, about 45% were eligible for free or reduced-price lunch, about 24% reported that their primary language was not English, and about 4% were identified as homeless. Table 1 provides summary statistics of school-level demographic characteristics.

Measures

We look at four broad categories of performance measures: student academic performance, student behavior, racial/ethnic disparities in school performance, and school internal processes. Each of these categories includes a number of sub-measures.

Student academic performance. Measures of student academic performance include *average math and reading scores* from state assessments, standardized within a given grade, year, and test, along with the *proportion of students who achieve proficiency in math and reading*. *School growth in math and reading* is calculated in terms of school value-added and median student growth percentile. School value-added captures the degree to which learning gains during the year are greater than the estimated gains predicted from students' prior achievement, student characteristics, and school demographics and prior performance. Median student growth percentile is a more commonly used measure for assessing growth in states' ESSA plans. Appendix A describes our model specifications for both value-added and median student growth percentile. Finally, we look at the *five-year adjusted cohort graduation rate for high schools*, which is the percentage of students from a ninth-grade cohort who graduate by their fifth year at the school, adjusting for transfers in and out of state public schools.

Student behavior. We look at two measures of student behavior. *Chronic absenteeism rate* is the percentage of students at a school who have ≥ 18 days of excused or unexcused absences during the school year, which is equivalent to 10% of a school year. *Disciplinary infraction rate* is the percentage of students at a school with at least one formal disciplinary infraction during the school year.

Racial/ethnic disparities. We measure performance disparities for racial and ethnic groups of students in two categories:

1. *Historically underserved students of color* (hereafter referred to as HUSC), which includes American Indian/Alaska Native, Black/African American, Hispanic/Latino, and Pacific Islander students; and
2. *Non-HUSC*, which includes all remaining groups within the seven federally defined racial/ethnic categories—White, Asian, and multiracial students.

Although our definition of HUSC may not be consistent with that used in some prior studies, it is derived from achievement data from Washington State. The four groups we identify as HUSC all performed below the state average on both math and reading tests, while the three remaining groups performed above average.

In most cases, we define racial/ethnic disparity as the difference in means (or medians for student growth percentile) for these two groups of students. However, we use a different measure when comparing their scale score performance. For these performance measures, we calculate the *V-disparity*, separately for math and reading. The V-disparity (or V-gap, as Ho and Reardon [2012] call it), can be understood as the difference in mean test scores between two groups, both with standard normal test score distributions. This measure allows us to compare test score disparities across different test scales. We calculate disparities when there are at least 10 students in each of the two groups (HUSC and non-HUSC).²

² We set this minimum subgroup size to ensure that the measures are reliable and accurate representations of school disparities without excluding schools that have smaller yet still important numbers of students in particular subgroups. Given that chronic absenteeism and disciplinary infractions are relatively rare, the risk differences are relatively small. Compared with another measure of disparity in the literature—risk ratio (i.e., the HUSC rate divided by the non-HUSC rate)—risk difference might underestimate disparities. However, risk ratio has an inherent problem in that as one group's risk approaches 0.5, the maximum risk ratio is 2. This generates statistical artifacts, prohibiting us from observing the underlying relationship between average performance and disparities. The risk difference measure is also more consistent with the way in which many states display chronic absenteeism rates on their websites. States often show the rates among subgroups of students side-by-side; our approach of taking the differences between HUSC and non-HUSC peers conveys a similar meaning.

School internal processes. To capture school conditions and supports for learning, we identified common items in Seattle Public Schools' student and staff surveys that best capture educational practices and conditions that are malleable to school policies and practices. Using exploratory principal component factor analysis and orthogonal varimax (Kaiser off) rotation, we developed four composite constructs: *positive classroom peer environment*, or whether teachers manage their classrooms to create an inclusive and friendly environment focused on learning; *safe and welcoming school environment*, or whether the school creates a sense of pride, respect, support, and safety among students; *effective school management and procedures*, or whether the school engages teachers in decision-making, promptly and effectively resolves conflicts among staff, and uses consistent processes for supporting students who struggle; and *teaching and teacher supports*, or whether the school supports collaboration among faculty members and provides other supports for effective instruction. A full list of survey items can be found in Appendix B.

Analysis and Results

Below, for each type of measure we look at in relation to school performance, we introduce the analytic methods we used and summarize our key findings.

Student Behavioral and Academic Measures and School Performance

We examine the relationships between measures of school performance by calculating correlation coefficients for each pair of performance measures. We separate the measures into two groups—elementary and middle schools (grades 3–8) and high schools (grades 10–11)—based on the accountability system designs that most states and districts use and the availability of measures at different educational levels. Because the range of some measures (e.g., chronic absenteeism rates and disciplinary infraction rates) is relatively restricted, which can attenuate correlation coefficients, we use scatterplots to visually inspect the relationships between these measures. The scatterplots suggest patterns that are consistent with those reflected by the correlation coefficients. Since coefficients provide a more succinct way of summarizing the patterns than scatterplots do, we present mainly coefficients in this section and include the scatterplots in Appendix B. We then present an exploratory factor analysis of these performance measures to assess the dimensionality of the data.

Elementary and middle schools. As shown in Panel A of Table 2, for elementary and middle schools, average test scores for math and reading are highly correlated with one another ($\rho = .91$). The proficiency levels for math and reading (i.e., the percentage of students meeting standards on math and reading exams) are also highly correlated ($\rho = .87$). When looking across measurement types for the same subject (e.g., the average test score for math and the proficiency rate for math), we observe similarly high correlation coefficients (e.g., $\rho = .93$ or $.90$).

In addition to performance levels, we look at two types of growth measures: school value-added and median student growth percentile. The growth measures captured by school value-added are positively correlated with achievement levels. However, the correlation coefficients are relatively small, ranging from .10 to .28. In contrast, median student growth percentile has much greater correlations with average school performance levels in terms of both test scores and proficiency ($\rho = .34$ – $.56$). The reason for this difference in correlations may be that schools' average performance levels are influenced by student demographics and socioeconomic status, whereas the value-added growth measures deliberately account for the influence of these student characteristics and school contextual factors. In addition, while schools' average performance and growth (value-added or student growth percentile) are positively correlated (i.e., schools with high average scores tend to

have high growth), the low to moderate size of the coefficients suggests they do not always align with one another.

Table 2

Correlations Between Student Academic and Behavioral Measures in Washington State, 2010–2016

Panel A. Elementary and Middle Schools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Average math score	1.00									
(2) Average reading score	.91	1.00								
(3) % proficient in math	.93	.85	1.00							
(4) % proficient in reading	.81	.90	.87	1.00						
(5) Growth in math (VA)	.26	.15	.22	.10	1.00					
(6) Growth in reading (VA)	.17	.28	.14	.20	.57	1.00				
(7) Growth in math (SGP)	.53	.39	.51	.34	.81	.42	1.00			
(8) Growth in reading (SGP)	.49	.56	.47	.50	.40	.73	.53	1.00		
(9) Chronic absenteeism rate	-.36	-.36	-.40	-.37	-.08	-.08	-.18	-.20	1.00	
(10) Disciplinary infraction rate	-.37	-.39	-.41	-.40	-.10	-.13	-.19	-.22	.53	1.00

Panel B. High Schools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Average math score	1.00						
(2) Average reading score	.66	1.00					
(3) % proficient in math	.69	.52	1.00				
(4) % proficient in reading	.47	.51	.46	1.00			
(5) Graduation rate	.56	.58	.41	.46	1.00		
(6) Chronic absenteeism rate	-.49	-.39	-.44	-.38	-.31	1.00	
(7) Disciplinary infraction rate	-.37	-.42	-.35	-.21	-.36	.37	1.00

Note. VA = value-added. SGP = student growth percentile.

Moreover, the two behavioral measures—chronic absenteeism rates and disciplinary infraction rates—are closely related ($\rho = .53$). These two measures are also negatively correlated with the measures of academic performance (i.e., math and reading performance levels and growth) as expected, although the correlation coefficients are low to moderate ($\rho = -.08$ to $-.41$).

Table 3
Principal Factor Loadings of All School Performance Measures

Panel A. Elementary and Middle Schools

	Factor 1	Factor 2	Factor 3	Factor 4
(1) Average math score	.94			
(2) Average reading score	.94			
(3) % proficient in math	.93			
(4) % proficient in reading	.90			
(5) Growth in math (VA)		.87		
(6) Growth in reading (VA)			.82	
(7) Growth in math (SGP)		.81		
(8) Growth in reading (SGP)			.71	
(9) Chronic absenteeism rate				.52
(10) Disciplinary infraction rate				.52

Panel B. High Schools

	Factor 1	Factor 2	Factor 3
(1) Average math score	.89		
(2) Average reading score	.55		
(3) % proficient in math	.89		
(4) % proficient in reading	.42		
(5) Graduation rate		.45	
(6) Chronic absenteeism rate			.57
(7) Disciplinary infraction rate			.56

Note. Factor loadings $< .4$ are omitted. VA = value-added. SGP = student growth percentile.

High schools. As shown in Panel B of Table 2, for high schools, the correlation between test scores in math and reading is lower than the same correlation for elementary and middle schools ($\rho = .66$ vs. $.91$). We observe similar positive correlations between scale scores and proficiency levels, although again with coefficients lower than the corresponding ones for elementary and middle schools. We also see that all test score measures are positively correlated with five-year graduation rates ($\rho = .41$ – $.58$). The two behavioral measures, chronic absenteeism rates and disciplinary infraction rates, are negatively correlated with academic performance as measured by test scores and graduation rates, with correlation coefficients comparable to those calculated for elementary and middle schools. The two behavioral measures are also positively correlated with one another ($\rho = .37$).

To further explore the multidimensionality of these measures, we use exploratory principal component factor analysis and orthogonal varimax (Kaiser off) rotation. As shown in Table 3, four factors are discovered for elementary and middle schools, with Factor 1 focusing on student achievement levels, Factor 2 on math growth, Factor 3 on reading growth, and Factor 4 on student behavior. We observe three factors for high schools, with Factor 1 focusing on student achievement levels, Factor 2 on graduation rates, and Factor 3 on student behavior. These factor loading patterns may largely be consistent with the original intent of the policy design—namely, to include measures that go beyond achievement levels and capture other school outputs that stakeholders value.

Racial/Ethnic Disparities in Student Behavioral and Academic Measures and School Performance

Under ESSA, states are required to assess school performance using the whole school as the unit of analysis and by student subgroup. Thus, we ask: Are schools that are higher performing on average equally effective for HUSC and non-HUSC? To answer this question, we compare measures of average school performance with measures of the disparities in schools' contributions to HUSC and non-HUSC. We start the analysis by presenting simple scatterplots showing the performance of HUSC and non-HUSC in schools with varying average school performance. We further explore the relationship between average performance and HUSC disparities by running regression models that control for other school characteristics (e.g., percentage of HUSC, percentage eligible for free or reduced-price lunch, log enrollment, percentage whose primary language is English, percentage of homeless students, percentage of female students) and Washington State time trends. Standard errors are clustered at the school level.

Elementary and middle schools. Panels A and B of Figure 1 show math and reading scale scores for elementary and middle schools. As expected, we observe a positive relationship between average school performance and the average performance of both HUSC and non-HUSC. However, the slopes of the lines appear to diverge between these two groups in the scatterplot for math scores, with a steeper positive slope for non-HUSC than for HUSC. In other words, while HUSC and non-HUSC both perform better in higher performing schools on average, HUSC tend to fall further behind non-HUSC in math as school performance increases. Therefore, a disparity measure would be larger in high-performing schools than in low-performing schools. It is not necessarily the case that high-performing schools on average are less able to support the learning of HUSC than other schools are, because HUSC in these schools, on average, still tend to be higher performing than HUSC in other schools; rather, this correlation suggests that HUSC do not reach the same level as their non-HUSC peers in the same high-performing schools.

Panels C and D present the relationships for math and reading proficiency rates. For both subjects, we see positive trends for both HUSC and non-HUSC, and the schools with the highest proficiency rates (at the far right of the graph) tend to have narrower gaps between the two groups.

Given that attaining proficiency requires a student to reach a certain set bar, it is not that surprising that the disparities in high-proficiency schools are smaller. Once a school reaches a certain overall proficiency level, there is less room for disparities in proficiency rates to exist. However, given the wider disparities in average scale scores at higher performing schools, it seems that proficiency, which is a coarse measure relative to scale score, may mask some of the differences in achievement between student subgroups within schools (Polikoff, Duque, & Wrabel, 2016).

The regression results in Table 4 formalize the patterns described above. V-disparity, which indicates HUSC minus non-HUSC test scores, increases (becomes more negative) as average school test scores increase. For a one standard deviation increase in average school test scores, HUSC fall about 10% or 13% of a standard deviation further behind their non-HUSC peers in the same school. Moreover, the associations between average school proficiency rate and proficiency rate differences are statistically nonsignificant.

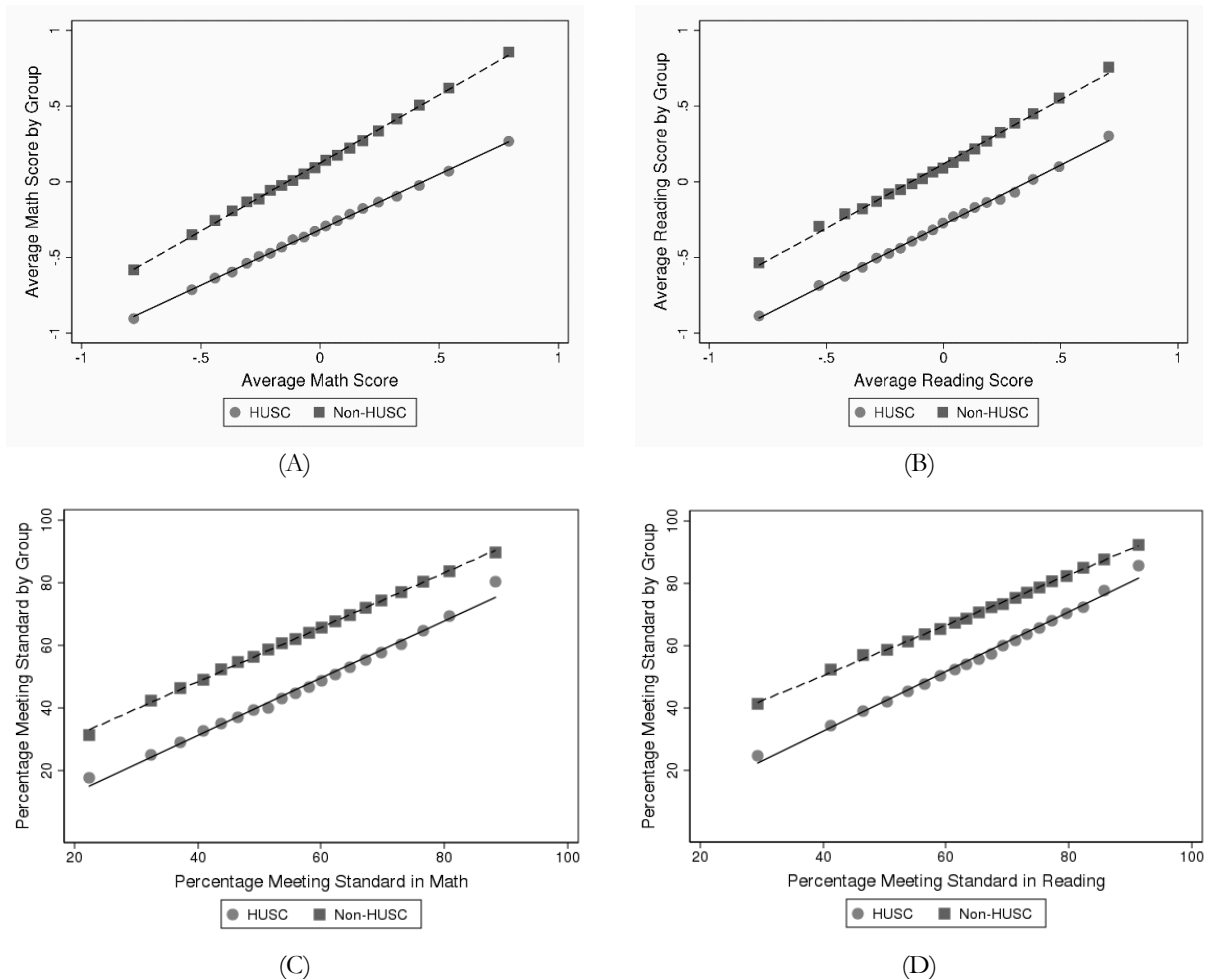


Figure 1. Student academic achievement in elementary and middle schools: Relationships between average school performance and HUSC disparities in Washington State, 2010–2016

Note. To enable the clear observation of relationships, the scatterplots consolidate raw information about all Washington State public schools over six years into 20 equal-sized bins. This reduces the clutter of dots on the scatterplots and makes patterns easier to discern visually.

Table 4
Regression of the Relationship Between Average School Performance and HUSC Disparities for Elementary and Middle Schools in Washington State, 2010–2016

	V-Disparity Scale Scores		Difference in Proficiency Rates		Value-Added Disparity		SGP Disparity		Risk Difference	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Chronic Absences	Disciplinary Infractions
Avg. school performance	-0.103** (0.021)	-0.131** (0.023)	-0.019 (0.018)	0.038 (0.021)	-0.007 (0.012)	0.030* (0.015)	-0.032** (0.011)	-0.022 (0.015)	0.102* (0.042)	0.197** (0.050)
N	9,573	9,598	9,628	9,633	7,046	7,046	6,388	6,391	2,338	937

Note. SGP = student growth percentile. Standard errors clustered by school are shown in parentheses. V-disparity measures the gap in scale scores between HUSC and non-HUSC. Difference in proficiency is the difference in proficiency rates between HUSC and non-HUSC. Value-added disparity is equal to the difference between school value-added measures for HUSC and non-HUSC. SGP disparity is the difference between median SGP for HUSC and non-HUSC at a given school. Risk differences in chronic absenteeism and disciplinary infractions are calculated using HUSC risks minus non-HUSC risks. We set a minimum n of 10 (i.e., at least 10 students need to have observed data for each group at the school). Value-added is estimated starting in 2010–11 and only for grades 3–8. All models include school-level covariates and year fixed effects. Full model results are presented in Appendix Table B2 (academic outcomes) and Appendix Table B4 (behavioral measures).

* $p < .05$. ** $p < .01$.

Panels A and B of Figure 2 plot trends for HUSC and non-HUSC in math and reading value-added. Value-added controls not only for students' prior test scores but also for several student and school characteristics, such as whether the student is eligible for free or reduced-price lunch, disabled, or homeless, as well as similar student population characteristics at the school level. Given that these controls are likely correlated with race/ethnicity, it is unsurprising that the gaps in value-added are very narrow. However, we find an interesting pattern in reading value-added. The trend lines seem to indicate that the disparities between HUSC and non-HUSC in value-added decline as the school's overall value-added increases. This finding suggests, at least in reading, that higher value-added schools are more effective than lower value-added schools at reducing equity gaps.

Panels C and D present the scatterplots for another measure of student growth—median student growth percentile. This measure does not control for student and school characteristics as value-added does and is therefore still correlated with these characteristics. Most likely due to this difference, the patterns differ from the ones we see in Panels A and B. For both math and reading, schools with lower median student growth percentiles have smaller differences between HUSC and non-HUSC, while schools with higher median student growth percentiles have larger disparities.

The relationships observed in Figure 2 are further confirmed in our regression analyses. Table 4 shows a statistically nonsignificant relationship between math value-added and racial disparities in math value-added, but it shows a significant positive relationship between reading value-added and racial disparities in reading value-added. A one standard deviation increase in average value-added at a school is correlated with a 3% of one standard deviation decrease in the difference in value-added, suggesting that high-growth schools in reading tend to be more equal schools. In terms of median student growth percentile, however, the relationship between school average and disparities is reversed. With a 1 percentile increase in a school's median student growth percentile, there is about a 3% increase in the differences in median math student growth percentiles between HUSC and non-HUSC groups in the same school. The relationship between reading value-added and racial disparities in reading value-added is nonsignificant.

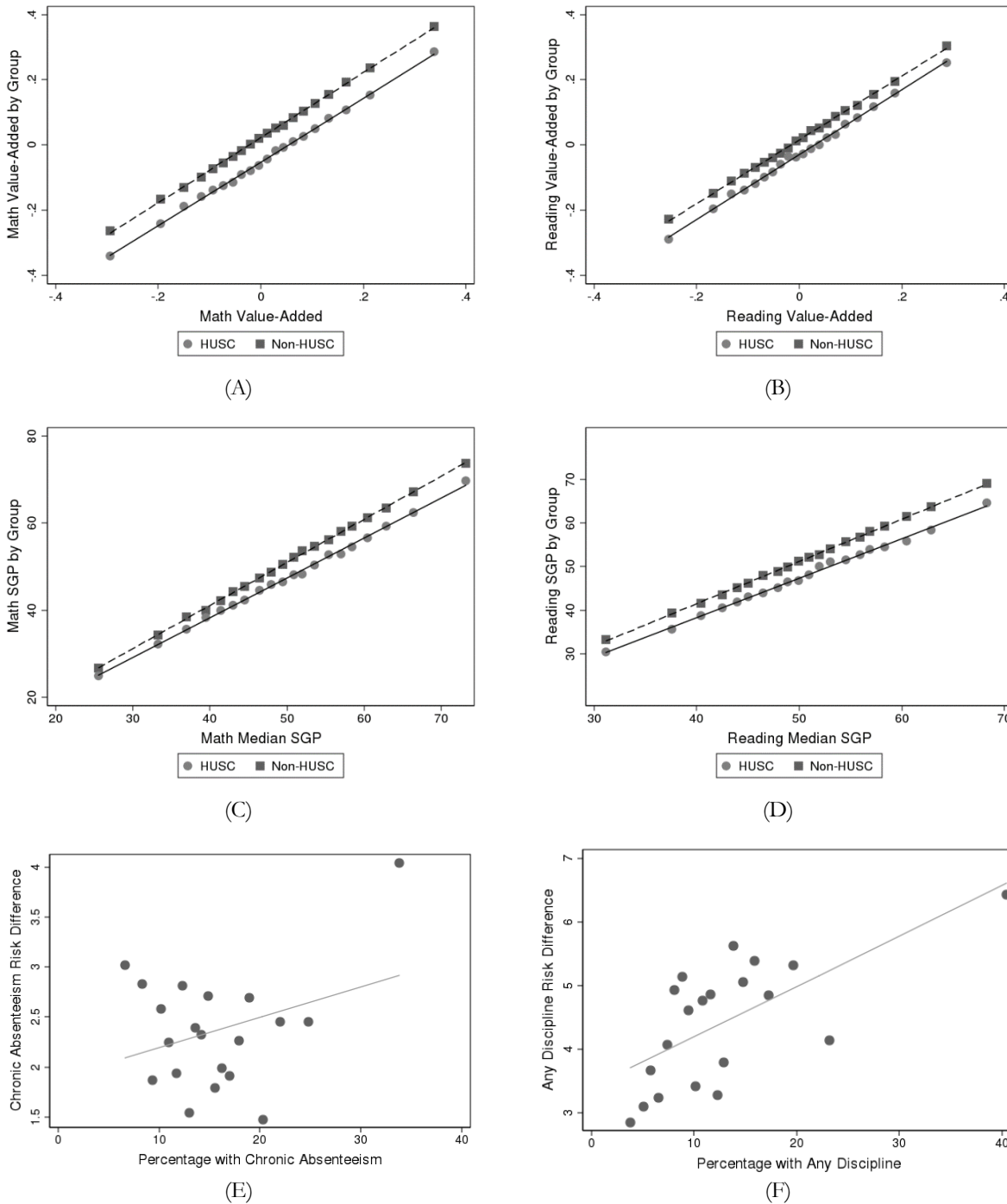


Figure 2. Student academic growth in elementary and middle schools: Relationships between average school performance and HUSC disparities in Washington State, 2010–2016
Note. To enable the clear observation of relationships, the scatterplots consolidate raw information about all Washington State public schools over six years into 20 equal-sized bins. This reduces the clutter of dots on the scatterplots and makes patterns easier to discern visually.

We next explore how schools’ performance in behavioral outcomes compares with HUSC disparities in these same measures. Panels E and F of Figure 2 reveal a positive relationship between the average school levels and risk differences for chronic absenteeism and disciplinary infractions

(i.e., the HUSC rate compared with the non-HUSC rate). Schools with higher average chronic absenteeism or disciplinary infraction rates tend to show higher disparities, with greater risk among HUSC than among their non-HUSC peers. The last two columns in Table 4 confirm the pattern shown in the scatterplots, even after controlling for school characteristics and state time trends.³ A 10 percentage point increase in chronic absenteeism rates (or disciplinary infraction rates) at a school would increase the risk difference at the average elementary or middle school by 1 percentage point (2 percentage points for disciplinary infractions).

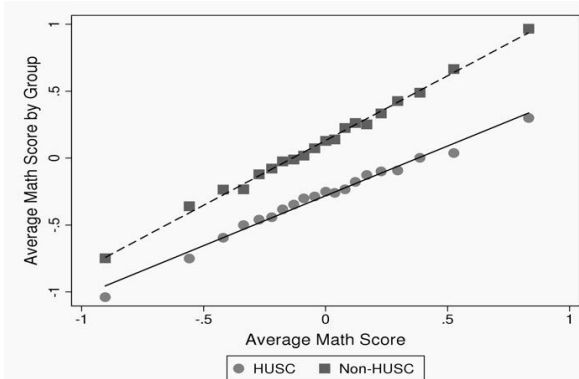
High schools. As shown in Panel A of Figure 3, which plots HUSC and non-HUSC average math test scores (for 10th and 11th graders), we see a more pronounced divergence in the trend lines than we do for elementary and middle schools (Figure 1, Panel A). That is, we see a more obvious difference in the disparities between high- and low-performing schools for math test scores. Again, the trend lines for reading scores show no patterns of divergence or convergence (Panel B). Table 5 presents the estimated coefficients for the regression of high school disparity measures on average school performance and shows that for a one standard deviation increase in average test scores, there would be a 26% of one standard deviation increase in disparities in math and a 16% of one standard deviation increase in disparities in reading.

Panels C and D of Figure 3 present the analogues to the scatterplots shown in Panels C and D of Figure 1 for proficiency levels. For math, the lines diverge greatly in the highest performing schools, with HUSC falling further behind their non-HUSC peers in the same school as average school performance increases. This is in contrast with Panel C of Figure 1, which shows almost parallel lines with a slight narrowing of disparities in the highest performing schools. We see a similar pattern for reading proficiency in high schools (10th or 11th grades) as we do in elementary and middle schools. The lines for these two groups are almost parallel, although in the highest performing high schools, the lines converge, suggesting a reduction in gaps. The regression results in Table 5 show similar patterns.

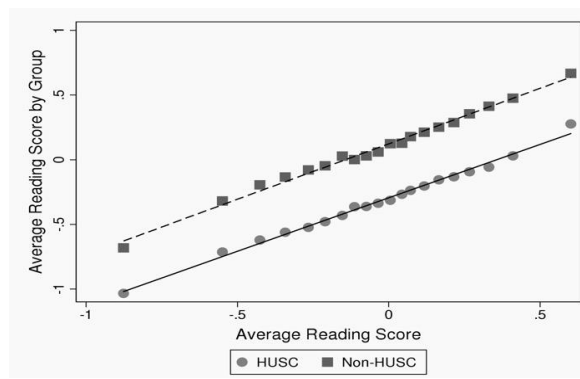
Panel E of Figure 3 plots the graduation rates for HUSC and non-HUSC. The disparities in graduation rates appear to narrow as the school's overall graduation rate increases. This makes sense given that graduation provides a set bar for students to cross. In high-performing schools overall, both HUSC and non-HUSC are likely to reach this bar; we thus see a converging trend between HUSC and non-HUSC when measuring their performance in terms of graduation rates. However, after controlling for school-level factors, the regression results in Table 5 show a statistically nonsignificant narrowing of the graduation rate disparity.

Lastly, Panels F and G of Figure 3 reveal a largely positive relationship between average levels and risk differences for chronic absenteeism and disciplinary infractions, similar to the patterns seen in elementary and middle schools. Schools with higher average chronic absenteeism rates (or disciplinary infraction rates) tend to show higher disparities between HUSC and non-HUSC. The last two columns in Table 5 confirm these observations even after controlling for school characteristics and state time trends. A 10 percentage point increase in chronic absenteeism rates (or disciplinary infraction rates) at an average high school would increase risk differences by about 3 percentage points.

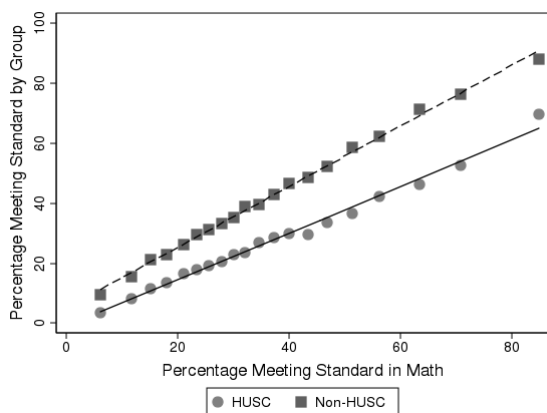
³ One could argue that these performance measures are a result of district-level attendance/discipline policies and that therefore it would be appropriate to include district-level effects. The analyses including district-level effects yield findings similar to those presented in this paper. Results are available upon request.



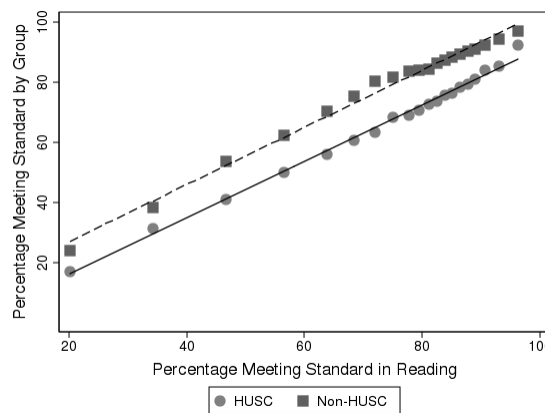
(A)



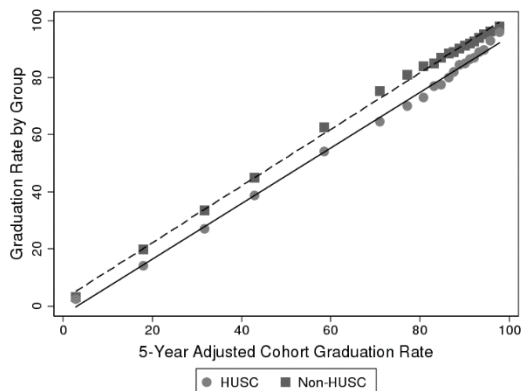
(B)



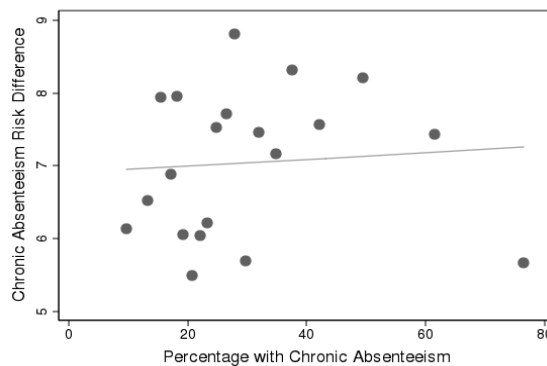
(C)



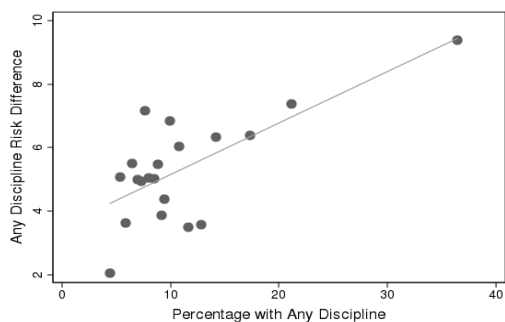
(D)



(E)



(F)



(G)

Figure 3. Student academic achievement in high schools: Relationships between average school performance and HUSC disparities in Washington State, 2010–2016

Note. To enable the clear observation of relationships, the scatterplots consolidate the raw information about all Washington State public schools over six years into 20 equal-sized bins. This reduces the clutter of dots on the scatterplots and makes patterns easier to discern visually.

Table 5
Regression of the Relationship Between Average School Performance and HUSC Disparities for High Schools in Washington State, 2010–2016

	V-Disparity Scale Score		Difference in Proficiency Rates		Difference in Graduation Rates	Risk Difference	
	Math	Reading	Math	Reading		Chronic Absences	Disciplinary Infractions
Average school performance	-0.258** (0.044)	-0.157** (0.060)	-0.188** (0.036)	-0.040 (0.042)	0.000 (0.000)	0.029 (0.032)	0.340** (0.077)
N	1,334	1,677	1,423	1,712	1,856	740	194

Note. Standard errors clustered by school are shown in parentheses. V-disparity measures the achievement gap between HUSC and non-HUSC. Difference in proficiency/graduation is the difference in proficiency/graduation rates between HUSC and non-HUSC. We set a minimum n of 10 (i.e., at least 10 students need to have observed data for each group at the school). Graduation rates are available starting in 2010–11. Risk differences in chronic absenteeism and disciplinary infractions are calculated using HUSC risks minus non-HUSC risks. All models include school-level covariates and year fixed effects. Full model results are presented in Appendix Table B3 (academic measures) and Appendix Table B5 (behavioral measures).

* $p < .05$. ** $p < .01$.

School Process Measures and School Performance

Our final research question relates to how well school processes—as captured by student and staff satisfaction surveys—predict schools’ academic performance and disparities between HUSC and non-HUSC.⁴ Using data from Seattle Public Schools, we regress each measure of schools’ average performance and disparities on the composites of process measures by controlling for a lagged dependent variable in each regression, school covariates, and year fixed effects. Table 6 presents the main results, with Panel A focusing on average school performance and Panel B focusing on disparities.⁵

Coefficients for all but two of the measures are statistically nonsignificant. First, more positive staff perceptions of instructional supports appear to be positively associated with higher average test scores and proficiency rates in both math and reading. Second, more positive student perceptions of the school environment (i.e., whether it is safe and welcoming) appear to be associated with smaller differences in reading growth (student growth percentile) between HUSC and non-HUSC. The overall weak correlations between these process measures and student achievement outcomes may be indicative of the multidimensionality of school performance, as many school actions may not result in immediate impacts on test scores. Alternatively, these weak relationships may be due to this study’s data limitations. We have a relatively small sample size, and survey measures may contain measurement error, both of which reduce the study’s analytical power.

⁴ We did not estimate the association between school process measures and chronic absenteeism rates, disciplinary infraction rates, or graduation rates because we do not have sufficient sample sizes. We have only about 30 schools—years that have both process and outcome information available.

⁵ Unlike other results presented in this manuscript, we show the results for elementary, middle, and high schools together. We do this because we only have a small number of high schools in the Seattle Public Schools (17). However, the results shown in Table 6 are consistent with those including only elementary and middle schools.

Table 6
Regressions of School Process Composites on School Performance and HUSC Disparities in Seattle Public Schools, 2014–2016

Panel A. Predicting Average School Performance

	Avg. Score		Proficiency		Growth (VA)		Growth (SGP)	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Measures From Student Surveys								
Positive classroom peer environment	0.024 (0.057)	0.110 (0.073)	2.246 (2.208)	4.567 (2.851)	-0.007 (0.060)	0.024 (0.054)	1.601 (5.010)	0.020 (4.579)
Safe and welcoming school environment	-0.105 (0.119)	-0.137 (0.105)	-2.808 (4.289)	-1.692 (3.121)	-0.029 (0.118)	-0.006 (0.100)	5.939 (9.817)	3.451 (6.480)
Measures From Staff Surveys								
Effective school management and procedures	0.053 (0.034)	0.032 (0.032)	2.119 (1.299)	0.627 (1.239)	0.022 (0.031)	0.003 (0.026)	4.735 (2.382)	1.703 (2.119)
Teaching and teacher supports	0.142** (0.052)	0.124* (0.054)	5.793* (2.218)	4.787** (1.498)	0.022 (0.033)	0.006 (0.029)	-0.917 (5.254)	-2.684 (3.082)
N	229	238	229	238	203	203	75	75

Panel B. Predicting Disparities

	V-Disparities in Average Score		Proficiency Difference		Growth Difference (VA)		Growth Difference (SGP)	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Measures From Student Surveys								
Positive classroom peer environment	0.127 (0.087)	0.089 (0.086)	2.266 (3.475)	3.018 (3.193)	0.071 (0.056)	0.108 (0.059)	1.219 (5.244)	0.773 (5.807)
Safe and welcoming school environment	0.130 (0.120)	-0.018 (0.131)	-2.044 (5.170)	1.202 (4.196)	0.003 (0.076)	-0.014 (0.095)	-4.460 (7.516)	-27.402* (11.886)
Measures From Staff Surveys								
Effective school management and procedures	0.028 (0.032)	0.040 (0.039)	2.471 (1.402)	1.720 (1.418)	-0.007 (0.020)	-0.022 (0.030)	-1.895 (2.185)	2.045 (2.935)
Teaching and teacher supports	0.002 (0.050)	0.076 (0.054)	-0.371 (1.974)	0.433 (1.810)	0.050 (0.033)	0.006 (0.032)	1.969 (3.573)	-6.538 (5.186)
N	215	224	218	224	191	190	71	70

Note. VA = value-added. SGP = student growth percentile. Standard errors clustered by school are shown in parentheses. Each regression includes the lagged dependent variable, school-level covariates, and year fixed effects. Full model results are available from the authors upon request. * $p < .05$. ** $p < .01$.

Discussion

Contributions to the Literature

Although prior studies have illustrated the difficulty of using multiple measures to consistently identify schools in need of improvement (Hough et al., 2016), they have not conceptualized these issues in a coherent framework. This study, to our knowledge, is the first to conceptualize the sources of the multidimensionality of school performance by drawing on both management science and education literature, which include stakeholders' values and interests in education, strategic choices regarding initiatives and resource allocation, interactions between school services and student backgrounds, and the temporal relationship between organizational activities and outcomes. Moreover, we empirically illustrate three key school performance measures as the manifestation of the inherent sources of organizational multidimensionality: first, that measures have different associations with one another and are loaded onto different factors; second, that school performance can be heterogeneous for different subgroups of students; and third, that the nature of the measures and the temporal relationships between them add another layer of complexity to the measurement of school performance.

In particular, our analyses of HUSC disparity measures are a key contribution to the literature on organizational multidimensionality and are not discussed in Richard et al.'s (2019) original framework. Richard et al.'s framework was developed for private companies; our conceptualization, meanwhile, foregrounds schools' mission of promoting educational equity and broader societal good. Our study also adds to the literature on educational inequality by inquiring into disparities in achievement and growth within schools. The negative relationships between average school performance and HUSC disparities in scale scores and median student growth percentiles suggest that HUSC in schools with high average performance were further behind their non-HUSC peers in the same school. In schools with higher rates of chronic absenteeism and disciplinary infractions, the relative risk for HUSC increases as well. In contrast, schools that promoted higher average growth for all students as captured by value-added measures shrank the gap between HUSC and non-HUSC in reading. Moreover, the relationship between average graduation rates and graduation disparities seems to suggest that schools that perform higher on average are more equal schools.

Although school production may explain some of the complex relationships between average school performance and within-school disparities, which suggests all schools need to do more to reduce disparities, the inconsistent patterns among measures may be more a reflection of the nature of the measures themselves. For example, measures of average performance based on scale scores and student growth percentile are often influenced by student socioeconomic and school contextual backgrounds, while value-added measures deliberately account for these factors. In other words, value-added measures level the playing field between HUSC and non-HUSC by eliminating the associations between student growth and student background characteristics (Ehlert et al., 2014). In addition, proficiency rates and graduation rates each have a set of criteria for all students to meet statewide, and these criteria tend to be basic. Once a school's average performance reaches a certain level, it is not hard for most students in the school, if not all students, to meet these criteria. We thus see smaller disparities in these measures between student subgroups in schools that are higher achieving on average. In sum, the psychometric attributes of the measures themselves add another layer of complexity to the measurement of school performance.

Practical Implications

At this early stage of ESSA implementation, virtually all states use a weighting system to combine scores from multiple measures into one index to differentiate between schools and identify underperforming schools for additional supports. Our findings suggest that besides this composite score, states should also inform the public of schools' performance on each measure individually. For example, Washington State has detailed weighting mechanisms that vary by grade level (Washington State Office of Superintendent of Public Instruction, 2018, p. 42). For K-12 schools with all indicators, English language arts and math proficiency are each worth 15%, growth is worth 25%, graduation 25%, English language progress 5%, and school quality or student success indicators (averaged) 15%. Delaware aggregates five indicators (academic achievement, academic progress, school quality/student success, graduation rate, and progress toward English language proficiency) to create a summative index score for schools, which is then translated into an overall categorical identification (e.g., exceeds, meets, or meets few expectations) (Delaware Department of Education, 2019, p. 49). New York State gives the greatest weight to academic achievement and growth in elementary and middle schools and academic achievement and graduation rates in high schools, followed by progress toward English language proficiency and then chronic absenteeism and a college and career readiness index (New York State Education Department, 2017, p. 70).

Although a composite index may provide a more holistic portrait of school performance than each single measure, it may mask important differences between schools on individual dimensions. Moreover, some measures (disciplinary infraction rates and chronic absenteeism rates) may be noisier than others (such as test scores or graduation rates). More complex school data profiles generated by multiple measures contain more noise or can fluctuate more dramatically from year to year. Another challenge for combining different measures into a single framework is to connect these measures to a coherent framework for student and school success. In the absence of a conceptual framework, school systems will simply be adding more measures to their accountability system. Given these concerns, it is important for states to publish not only the composite but also the individual measures. For example, Washington State school report cards display each measure for a given school or district and allow users to view the breakdown for each by student subgroup (Washington State Office of Superintendent of Public Instruction, 2019). This display allows users to interpret and select the information that is most consistent with their theory of action.

Second, our study highlights the importance of having a coherent school improvement framework to guide local districts and schools to carry out data-driven, evidence-based reform strategies. The appropriate use of data demands both technical expertise and a systemic understanding of the practical issues involved. For example, improving process measures such as collaborative teaching may predict student achievement gains, and policymakers and researchers can leverage the temporal relationship between these measures to diagnose issues and develop interventions to improve student outcomes. However, these process measures, often gathered through surveys or observations, contain measurement error, which may significantly attenuate the estimated relationships between school processes and outcome measures. What is more, the relationships between school processes and outcomes are inherently complex. School district leaders need a deep conceptual understanding of the mechanisms of change. This evidence-based practice demands high levels of both technical and theoretical expertise, which oftentimes practitioners may not have. Partnerships between researchers and practitioners that aim to build school districts' capacity for using data and evidence-based practices can be instrumental in this regard (Booker, Conaway, & Schwartz, 2019; Coburn & Penuel, 2016; Kane, 2017).

Limitations

We acknowledge several limitations of our work. First, this study by no means includes all measures that are used in school performance measurement systems across states under ESSA. Further research is needed to examine other popular measures (e.g., access to advanced coursework). Additionally, several measures—including student behavioral and school processes—may be noisy due to measurement error or underpowered due to small sample sizes. These issues limit this study's capacity to identify true relationships between these measures.

Conclusion

Despite these limitations, this paper provides useful empirical illustrations of the complex relationships between multiple measures used under the ESSA school performance measurement system. States and local practitioners will need consistent support as they select, collect data on, and use these measures to remedy underperformance and promote equity in schools.

Acknowledgements

We thank colleagues at the Office of Student Information of the Washington State Office of Superintendent of Public Instruction (OSPI) for granting the data access and for their useful feedback on an earlier draft. We particularly appreciate the support from Katie Weaver Randall and Julie Hoff. We also greatly appreciate the feedback on an earlier draft from Emily Penner and other APPAM participants.

References

- Bae, S. (2018). Redesigning systems of school accountability: A multiple measures approach to accountability and support. *Education Policy Analysis Archives*, 26(8), 1–32.
<https://dx.doi.org/10.14507/epaa.26.2920>
- Bentes, A. V., Carneiro, J., da Silva, J. F., & Kimura, H. (2012). Multidimensional assessment of organizational performance: Integrating BSC and AHP. *Journal of Business Research*, 65(12), 1790–1799. <https://doi.org/10.1016/j.jbusres.2011.10.039>
- Betebenner, D. W., Van Iwaarden, A., Domingue, B., & Shang, Y. (2014). SGP: An R package for the calculation and visualization of student growth percentiles & percentile growth trajectories. [Computer software manual, R package version 1.2-0.0].
<http://centerforassessment.github.com/SGP/>
- Bohanon, H., Fenning, P., Carney, K. L., Minnis-Kim, M. J., Anderson-Harriss, S., Moroz, K, B. ... Pigott, T. D. (2006). Schoolwide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions*, 8(3), 131–145.
<https://doi.org/10.1177/10983007060080030201>
- Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). *School composition and the Black–White achievement gap* (NCES 2015-018). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Booker, L., Conaway, C., & Schwartz, N. (2019). *Five ways RPPs can fail and how to avoid them: Applying conceptual frameworks to improve RPPs*. William T. Grant Foundation.
- Burris, C. C., Heubert, J. P., & Levin, H. M. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal*, 43(1), 137–154.
<https://doi.org/10.3102/00028312043001105>

- Callen, J. L. (1991). Data envelopment analysis: Partial survey and applications for management accounting. *Journal of Management Accounting Research*, 3(1), 35–56.
- Cameron, K. (1986). A study of organizational effectiveness and its predictors. *Management Science*, 32(1), 87–112. <https://doi.org/10.1287/mnsc.32.1.87>
- Carter, P. L. (2012). *Stubborn roots: Race, culture, and inequality in U.S. and South African schools*. Oxford University Press.
- Chakravarthy, B. S. (1986). Measuring strategic performance. *Strategic Management Journal*, 7(5), 437–458. <https://doi.org/10.1002/smj.4250070505>
- Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, 45(1), 48–54. <https://doi.org/10.3102/0013189X16631750>
- Crosnoe, R. (2009). Low-income students and the socioeconomic composition of public high schools. *American Sociological Review*, 74(5), 709–730. <https://doi.org/10.1177/000312240907400502>
- Darling-Hammond, L., & Adamson, F. (Eds.). (2014). *Beyond the bubble test: How performance assessments support 21st century learning*. Jossey-Bass.
- Delaware Department of Education. (2019). *Delaware's ESSA consolidated plan*. https://www.doe.k12.de.us/cms/lib/DE01922744/Centricity/Domain/208/DE_consolidatedstateplan.FINAL%20clean%2006.10.19.pdf
- Devinney, T. M., Yip, G. S., & Johnson, G. (2010). Using frontier analysis to evaluate company performance. *British Journal of Management*, 21(4), 921–938. <https://doi.org/10.1111/j.1467-8551.2009.00650.x>
- Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C. ... Giffin, J. (2017). *School improvement grants: Implementation and effectiveness* (NCEE 2017- 4013). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251. <https://doi.org/10.3102/0013189X15584327>
- Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. J. (2014). The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Statistics and Public Policy*, 1(1), 19–27. <https://doi.org/10.1080/2330443X.2013.856152>
- Flannery, K. B., Fenning P., Kato, M. M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly*, 29(2), 111–124. <http://dx.doi.org/10.1037/spq0000039>
- Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2015). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, 18(1), 41–51. <https://doi.org/10.1177/1098300715580992>
- Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, 44(5), 293–307. <https://doi.org/10.3102/0013189X15592622>
- Green, T. F. (1983). Excellence, equity, and equality. In L. S. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 318–341). Longman.

- Hallinan, M. T., & Williams, R. A. (1989). Interracial friendship choices in secondary schools. *American Sociological Review*, *54*(1), 67–78. <https://dx.doi.org/10.2307/2095662>
- Hamilton, L., Berends, M., & Stecher, B. (2005). *Teachers' responses to standards-based accountability*. RAND Education.
- Henri, J.-F. (2004). Performance measurement and organizational effectiveness: Bridging the gap. *Managerial Finance*, *30*(6), 93–123. <https://doi.org/10.1108/03074350410769137>
- Herman, R. D., & Renz, D. O. (2008). Advancing nonprofit organizational effectiveness research and theory: Nine theses. *Nonprofit Management and Leadership*, *18*(4), 399–415. <https://doi.org/10.1002/nml.195>
- Ho, A. D., & Reardon, S. F. (2012). Estimating achievement gaps from test scores reported in ordinal “proficiency” categories. *Journal of Educational and Behavioral Statistics*, *37*(4), 489–517. <https://doi.org/10.3102/1076998611411918>
- Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*(3), 133–144. <https://doi.org/10.1177/1098300709332067>
- Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children*, *70*(2), 185–199. <https://doi.org/10.1177/001440290407000204>
- Hough, H., Penner, E., & Witte, J. (2016). *Identity crisis: Multiple measures and the identification of schools under ESSA*. Stanford University, Policy Analysis for California Education.
- Ittner, C. D., Larcker, D. F., & Randall, T. (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*, *28*(7–8), 715–741. [https://doi.org/10.1016/S0361-3682\(03\)00033-3](https://doi.org/10.1016/S0361-3682(03)00033-3)
- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, *42*(6), 304–316. <https://doi.org/10.3102/0013189X13495087>
- Kalogrides, D., Loeb, S., & Bêteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, *86*(2), 103–123. <https://doi.org/10.1177/0038040712456555>
- Kane, T. J. (2017). Making evidence locally. *Education Next*, *17*(2), 52–58.
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Bill & Melinda Gates Foundation.
- Kolluri, S. (2018). Advanced Placement: The dual challenge of equal access and effectiveness. *Review of Educational Research*, *88*(5), 671–711. <https://doi.org/10.3102/0034654318787268>
- Lauen, D. L., & Gaddis, S. M. (2016). Accountability pressure, academic standards, and educational triage. *Educational Evaluation and Policy Analysis*, *38*(1), 127–147. <https://doi.org/10.3102/0162373715598577>
- Mickelson, R. A. (2003). When are racial disparities in education the result of racial discrimination? A social science perspective. *Teachers College Record*, *105*(6), 1052–1086. <http://dx.doi.org/10.1111/1467-9620.00277>
- Moody, J. (2002). Race, school integration, and friendship in America. *American Journal of Sociology*, *107*(3), 679–701. <https://doi.org/10.1086/338954>
- New York State Education Department. (2017). *New York's ESSA consolidated plan*. <http://www.nysed.gov/common/nysed/files/programs/essa/nys-essa-plan.pdf>
- Polikoff, M. S., Duque, M., & Wrabel, S. (2016). A proposal for school accountability under ESSA. <https://edexcellence.net/articles/top-essadesign-proposals-morgan-polikoff-matthew->

- duque-and-stephani-wrabel-university-of
- Raudenbush, S. W. (2009). Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings. *Education Finance and Policy*, 4(4), 468–491. <https://doi.org/10.1162/edfp.2009.4.4.468>
- Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(2), 40–68. <https://doi.org/10.7758/RSF.2019.5.2.03>
- Reardon, S. F., & Ho, A. D. (2015). Practical issues in estimating achievement gaps from coarsened data. *Journal of Educational and Behavioral Statistics*, 40(2), 158–189. <https://doi.org/10.3102/1076998615570944>
- Reardon, S., Kalogrides, D., & Shores, K. (2016). *The geography of racial/ethnic test score gaps* (CEPA Working Paper No. 16-10). Stanford University, Center for Education Policy Analysis.
- Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492–519. <https://doi.org/10.1162/edfp.2009.4.4.492>
- Reardon, S. F., & Robinson, J. P. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. A. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 497–516). Routledge.
- Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of Management*, 35(3), 718–804. <https://doi.org/10.1177/0149206308330560>
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal*, 42(1), 3–42. <https://doi.org/10.3102/00028312042001003>
- Sawhill, J. C., & Williamson, D. (2001). Mission impossible? Measuring success in nonprofit organizations. *Nonprofit Management and Leadership*, 11(3), 371–386. <https://doi.org/10.1002/nml.11309>
- Shores, K., Kim, H. E., & Still, M. (2017). *Categorical inequality in black and white: Linking disproportionality across multiple educational outcomes*. <https://ssrn.com/abstract=3013214>
- Skiba, R. J., Chung, C.-G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal*, 51(4), 640–670. <https://doi.org/10.3102/0002831214541670>
- Sparks, S. D. (2015, September 17). How does an equity audit work? [Blog post]. http://blogs.edweek.org/edweek/inside-school-research/2015/09/how_does_an_equity_audit_work.html
- Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing* (CSE Technical Report 482). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.
- Stosich, E. L., Bae, S., & Snyder, J. (2018). Redesigning assessment and accountability: An introduction. *Education Policy Analysis Archives*, 26(7), 1–10. <https://dx.doi.org/10.14507/epaa.26.3666>
- Sun, M., Liu, J., Zhu, J., & LeClair, Z. (2019). Using a text-as-data approach to understand reform processes: A deep exploration of school improvement strategies. *Educational Evaluation and Policy Analysis*, 41(4), 510–536.

- Sun, M., Penner, E. K., & Loeb, S. (2017). Resource-and approach-driven multidimensional change: Three-year effects of School Improvement Grants. *American Educational Research Journal*, 54(4), 607-643.
- Sun, M., Saultz, A., & Ye, Y. (2017). Federal policy and the teacher labor market: exploring the effects of NCLB school accountability on teacher turnover. *School Effectiveness and School Improvement*, 28(1), 102-122.
- Van der Stede, W., Chow, C. W., & Lin, T. W. (2006). Strategy, choice of performance measures, and performance. *Behavioral Research in Accounting*, 18(1), 185-205.
- Venkatraman, N., & Ramanujam, V. (1986). Measurement of business performance in strategy research: A comparison of approaches. *Academy of Management Review*, 11(4), 801-814. <https://doi.org/10.5465/amr.1986.4283976>
- Wallace, J. M., Jr., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, ethnic, and gender differences in school discipline among U.S. high school students: 1991-2005. *Negro Education Review*, 59(1-2), 47-62.
- Washington State Office of Superintendent of Public Instruction. (2018). *Washington's ESSA consolidated plan*. <https://www2.ed.gov/admins/lead/account/stateplan17/waconsolidatedstateplanfinal.pdf>
- Washington State Office of Superintendent of Public Instruction. (2019). *Washington State report card*. <https://washingtonstatereportcard.ospi.k12.wa.us/>
- West, M. R. (2016). Should non-cognitive skills be included in school accountability systems? Preliminary evidence from California's CORE districts. *Evidence Speaks Reports*, 1(13), 1-7.

About the Authors

Min Sun

University of Washington

misun@uw.edu

Min Sun (corresponding author) is an Associate Professor in Education Policy in the College of Education at the University of Washington. She specializes in economics of education, educator labor markets and quality, school accountability and improvement, and school finance policies.

Alec I. Kennedy

San Francisco Unified School District

kennedy@sfusd.edu

Alec Kennedy is currently an Educational Policy Analyst at San Francisco Unified School District. His work involves supporting the use of research evidence, conducting program evaluations using quantitative methods, and assisting in the review of data for continuous improvement. He received his PhD in Public Policy & Management from the University of Washington.

Eric M. Anderson

Seattle Public Schools

emanderson@seattleschools.org

Eric Anderson received his PhD in Political Science from Stanford University and currently is Director of Research & Evaluation for Seattle Public Schools. He leads a team of experienced researchers that support evidence-based strategic planning and continuous improvement efforts using performance analytics and mixed methods program evaluation research.

Appendix A: School-Level Student Growth Estimates

We calculate student growth for each school in two ways. First, we use a school value-added model, which captures the degree to which learning gains during the school year are greater than the estimated gains predicted based on students' prior achievement, student characteristics, and schools' average demographics and prior performance. We calculate school value-added using a two-step (averaged residuals) procedure, illustrated below (Ehlert et al., 2014).

$$\text{Step 1: } Y_{it} = \gamma_0 + Y_{i,t-1}\gamma_1 + X_{it}\gamma_2 + S_{it}\gamma_3 + u_{it}$$

$$\text{Step 2: } u_{it} = I_{it}\theta + e_{it}$$

Here, Y_{it} is the standardized test score for student i in school year t . Although a subject subscript is omitted, we calculate value-added separately for math and reading. $Y_{i,t-1}$ is a vector of prior-year test scores for both subjects for student i . X_{it} is a vector of student background characteristics that includes gender, grade level, eligibility for free or reduced-price lunch, homelessness, and an indicator of whether the student's primary language is English. S_{it} is a vector of school demographics that includes school-level aggregates of student background characteristics (e.g., percentage of female students, percentage of students eligible for free or reduced-price lunch, percentage of homeless students, and percentage of students whose primary language is English). Furthermore, S_{it} includes school-level racial demographics and prior average reading/math test scores. Once we estimate the first equation's regression model, we obtain residuals (u_{it}). In Step 2, we regress the residuals on a vector of indicators of student i 's school during year t (I_{it}). θ is a vector that contains the school value-added estimates.

Our second measure of growth uses a median *student growth percentile* among the students served by a given school. Like value-added, student growth percentile attempts to estimate the academic growth of a student compared with his or her peers. However, in contrast with value-added, student growth percentiles focus on a comparison group composed of students who are similar based only on prior academic performance. Student growth percentiles are estimated using quantile regression, which models quantiles of the outcome variable. Specifically, Washington State calculates student growth percentiles using the following method described by Betebenner and his colleagues (2014):

$$Q_{Y_t}(\tau|Y_{t-1}, \dots, Y_1) = \sum_{j=1}^{t-1} \sum_{i=1}^3 \phi_{ij}(Y_j)\beta_{ij}(\tau)$$

Here, $Q_{Y_t}(\tau|Y_{t-1}, \dots, Y_1)$ is the conditional quantile function using the τ -th conditional quantile of Y_t , the current year's test score. ϕ_{ij} denotes the B-spline basis function. The state calculates the student growth percentile using the student growth percentile package in R (Betebenner, Van Iwaarden, Domingue, & Shang, 2014). Student growth percentile is calculated for the individual student. To aggregate results to the school level, we calculate the commonly used median student growth percentile to assess growth.

Student growth percentiles are more descriptive and are not interpreted as showing the causal impacts of a school on an individual student's learning. Measures of school growth using student growth percentile are highly negatively correlated with the proportion of students in a school who are eligible for free or reduced-price lunch (Ehlert et al., 2014). In other words, economically disadvantaged schools often have lower student growth scores. In contrast, value-added measures provide a more causally sound and meaningful interpretation of schools' contributions to student learning.

Appendix B: Additional Figures and Tables

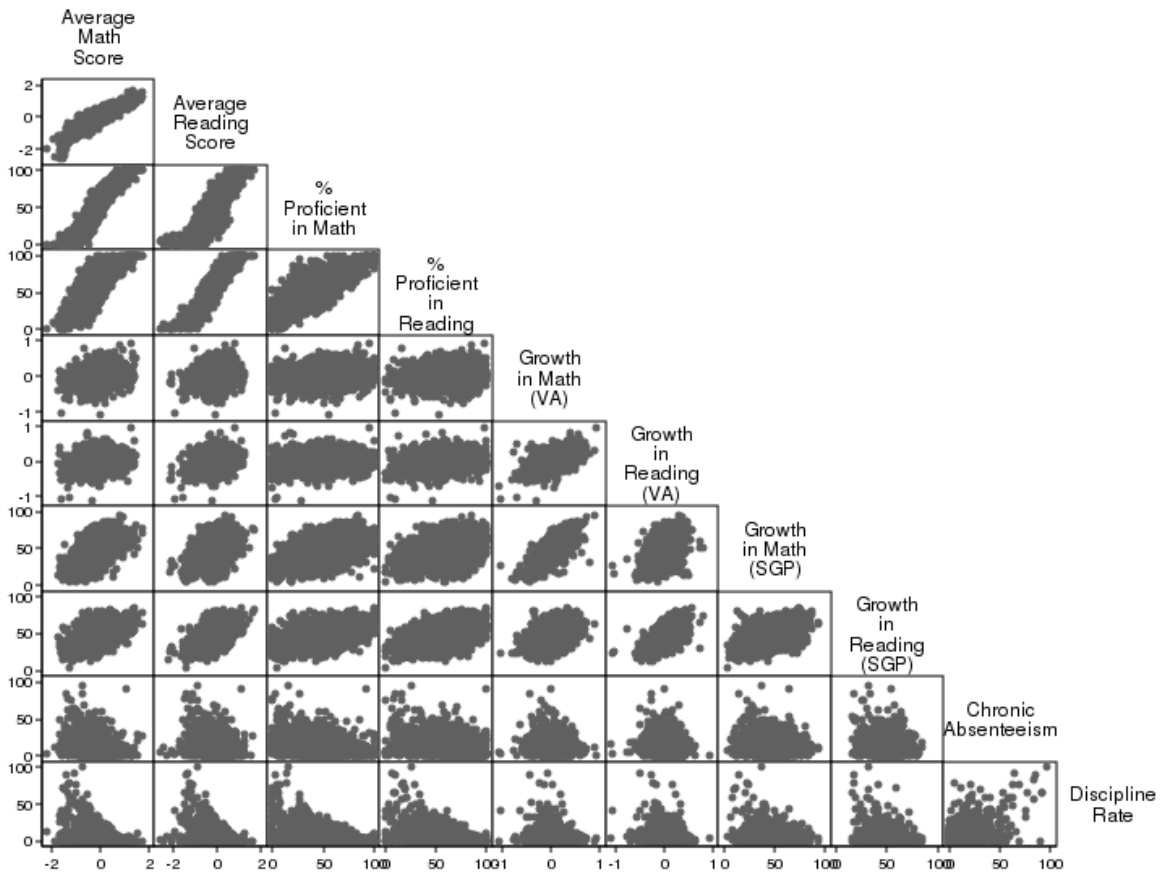


Figure B1. Scatterplots of relationships between multiple measures in elementary and middle schools
 Note. VA = value-added. SGP = student growth percentile.

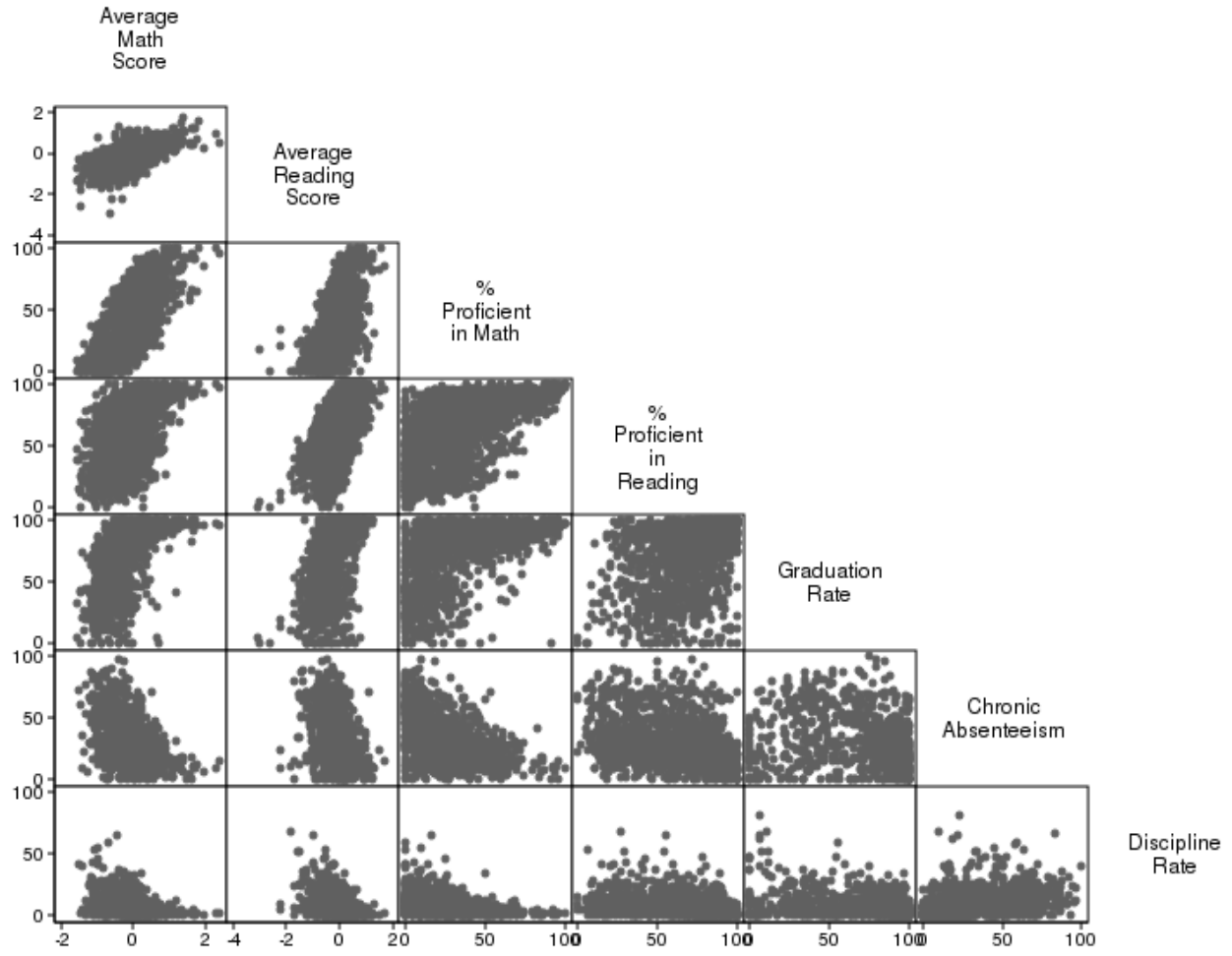


Figure B2. Scatterplots of relationships between multiple measures in high schools

Table B1

Seattle Public Schools Survey Items Used to Develop the Composite Measures of School Processes

Composite	Survey Questions
Positive classroom peer environment (Eigenvalue = 2.07, Cronbach's α = 0.73)	<ul style="list-style-type: none"> • Students in my class are focused on learning. • Students in my class are respectful to adults. • Students in my class help each other learn. • Students in my class are friendly to each other.
Safe and welcoming school environment (Eigenvalue = 8.3, Cronbach's α = 0.74)	<ul style="list-style-type: none"> • Adults at school care about me. • I enjoy going to school most days. • I feel safe in my classroom. • I am treated with as much respect as other students. • I feel proud of my school. • In my school, I feel that I belong to a group of friends.
Effective school management and procedures (Eigenvalue = 5.1, Cronbach's α = 0.84)	<ul style="list-style-type: none"> • This school has an effective process for making group decisions and solving problems. • Conflict among staff is resolved in a timely and effective manner. • I feel included in the decision-making process at this school. • This school has a consistent process for identifying students who struggle academically. • This school implements a clear plan of action when a student struggles academically. • I receive the support I need to address my students' behavior and discipline problems.
Teaching and teacher supports (Eigenvalue = 1.31, Cronbach's α = 0.78)	<ul style="list-style-type: none"> • This school meets regularly and often to discuss student data. • We share a common understanding of instructional best practices. • We have adequate time to plan collaboratively. • I have access to strategies and materials to support all learners in our classes. • I receive the support I need to differentiate and modify instruction for my students.

Table B2

Regression Coefficients for Relationships Between Average School Performance and HUSC Disparities for Elementary and Middle Schools

	V-Disparity		Difference		Value-Added Disparity		SGP Disparity	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Average performance	-0.103** (0.021)	-0.131** (0.023)						
Proportion proficient			-0.019 (0.018)	0.038 (0.021)				
Average growth					-0.007 (0.012)	0.030* (0.015)	-0.032** (0.011)	-0.022 (0.015)
% HUSC	-0.003** (0.001)	-0.004** (0.000)	-0.091** (0.019)	-0.129** (0.019)	-0.001** (0.000)	-0.001** (0.000)	-0.038** (0.012)	-0.044** (0.012)
% free/reduced-price lunch	0.001** (0.000)	0.001* (0.000)	0.062** (0.016)	0.071** (0.015)	0.001** (0.000)	0.001** (0.000)	0.051** (0.010)	0.032** (0.009)
% primary language is English	0.003** (0.000)	0.003** (0.000)	0.120** (0.018)	0.106** (0.017)	0.001** (0.000)	0.000** (0.000)	0.026** (0.010)	0.030** (0.009)
% homeless	-0.005* (0.002)	-0.005* (0.002)	-0.227** (0.075)	-0.189** (0.075)	-0.002* (0.001)	-0.002 (0.001)	-0.131* (0.062)	-0.112 (0.059)
% female	0.001 (0.002)	-0.001 (0.001)	-0.013 (0.057)	-0.075 (0.062)	0.001* (0.001)	-0.000 (0.001)	0.005 (0.044)	0.073 (0.042)
Log enrollment	0.013 (0.014)	0.033* (0.014)	-0.420 (0.542)	0.687 (0.537)	-0.008 (0.005)	0.010 (0.006)	0.248 (0.355)	-0.375 (0.339)
Middle school	-0.052** (0.012)	-0.038** (0.011)	-1.912** (0.477)	-1.264** (0.441)	-0.007 (0.004)	0.003 (0.004)	1.112** (0.259)	1.620** (0.256)
Constant	-0.798** (0.129)	-0.744** (0.127)	-22.732** (4.800)	-26.175** (4.944)	-0.143** (0.047)	-0.114* (0.054)	-7.071* (3.448)	-5.890 (3.225)
N	9,573	9,598	9,628	9,633	7,046	7,046	6,388	6,391

Note. SGP = student growth percentile. Standard errors clustered by school are shown in parentheses. V-disparity measures the achievement gap between HUSC and non-HUSC. Difference in proficiency is the difference in proficiency rates between HUSC and non-HUSC. Value-added disparity is equal to the difference between school value-added measures for HUSC and non-HUSC. SGP disparity is the difference between median SGP for HUSC and non-HUSC at a given school. We set a minimum cell size of 10 (i.e., at least 10 students need to have observed data for each group at the school). Value-added is estimated starting in 2010–11 and only for grades 3–8. All models also include year fixed effects.

* $p < .05$. ** $p < .01$.

Table B3

Regression of the Relationship Between Average Academic Effectiveness and HUSC Disparities for High Schools

	V-Disparity		Proficiency Difference		Grad. Rate Difference
	Math	Reading	Math	Reading	
Average performance	-0.258** (0.044)	-0.157** (0.060)			
Proportion proficient			-0.188** (0.036)	-0.040 (0.042)	
Graduation rate					0.000 (0.000)
% HUSC	-0.004** (0.001)	-0.005** (0.001)	-0.139** (0.049)	-0.127** (0.040)	0.001* (0.000)
% free/reduced-price lunch	0.001 (0.001)	0.003* (0.001)	0.080* (0.036)	0.066 (0.036)	-0.000 (0.000)
% primary language is English	0.004** (0.001)	0.003** (0.001)	0.146** (0.053)	0.084* (0.034)	0.001* (0.000)
% homeless	0.001 (0.007)	-0.002 (0.007)	-0.055 (0.132)	-0.163 (0.246)	0.000 (0.001)
% female	-0.005 (0.003)	-0.003 (0.004)	-0.084 (0.063)	-0.008 (0.085)	-0.000 (0.000)
Log enrollment	-0.038 (0.022)	0.002 (0.023)	-0.995 (0.682)	-0.158 (0.682)	-0.006 (0.005)
Constant	-0.621* (0.249)	-0.822* (0.320)	-21.627* (9.100)	-19.120 (10.315)	-0.072 (0.167)
<i>N</i>	1,334	1,677	1,423	1,712	1,856

Note. Standard errors clustered by school are shown in parentheses. V-disparity measures the achievement gap between HUSC and non-HUSC. Difference in proficiency/graduation is the difference in proficiency/graduation rates between HUSC and non-HUSC. We set a minimum cell size of 10 (i.e., at least 10 students need to have observed data for each group at the school). Graduation rates are available starting in 2010–11. All models include year fixed effects.

* $p < .05$. ** $p < .01$.

Table B4

Regression Coefficients of the Relationship Between Average School Performance in Behavioral Measures and HUSC Disparities for Elementary/Middle Schools

	Risk Differences in Chronic Absenteeism	Risk Differences in Disciplinary Infractions
School average risk	0.102* (0.042)	0.197** (0.050)
% HUSC	-0.105** (0.019)	-0.026 (0.029)
% free/reduced-price lunch	-0.007 (0.015)	-0.059** (0.021)
% primary language is English	0.018 (0.014)	-0.029 (0.025)
% homeless	-0.063 (0.075)	0.125 (0.103)
% female	-0.076 (0.055)	0.076 (0.097)
Log enrollment	0.307 (0.628)	2.167** (0.681)
Constant	6.573 (5.056)	-9.843 (6.633)
<i>N</i>	2,338	937

Note. Standard errors clustered by school are shown in parentheses. We set a minimum “cell” size of 10 (i.e., at least 10 students need to have a positive value for the event for each group at the school). Attendance and discipline data are only available starting in 2012–13. All models include year fixed effects.

* $p < .05$. ** $p < .01$.

Table B5

Regression of the Relationship Between Average School Effectiveness in Behavioral Measures and HUSC Disparities for High Schools

	Risk Differences in Chronic Absenteeism	Risk Differences in Disciplinary Infractions
School average risk	0.029 (0.032)	0.340** (0.077)
% HUSC	-0.084* (0.039)	-0.056 (0.050)
% free/reduced-price lunch	0.000 (0.028)	-0.133** (0.042)
% primary language is English	-0.092** (0.033)	-0.096* (0.042)
% homeless	-0.089 (0.067)	0.026 (0.219)
% female	0.043 (0.081)	0.173* (0.085)
Log enrollment	-0.086 (0.830)	-1.690 (1.465)
Constant	2.769 (7.755)	22.263 (13.445)
<i>N</i>	740	194

Note. Standard errors clustered by school are shown in parentheses. We set a minimum “cell” size of 10 (i.e., at least 10 students need to have a positive value for the event for each group at the school). Attendance and discipline data are only available starting in 2012–13. All models include year fixed effects.

* $p < .05$. ** $p < .01$.

education policy analysis archives

Volume 28 Number 89

June 8, 2020

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

Please send errata notes to Audrey Amrein-Beardsley at audrey.beardsley@asu.edu

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Lead Editor: **Audrey Amrein-Beardsley** (Arizona State University)

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Melanie Bertrand, David Carlson, Lauren Harris, Eugene Judson, Mirka Koro-Ljungberg, Daniel Liou, Scott Marley, Molly Ott, Iveta Silova** (Arizona State University)

Cristina Alfaro
San Diego State University

Gary Anderson
New York University

Michael W. Apple
University of Wisconsin, Madison

Jeff Bale
University of Toronto, Canada
Aaron Bevenot SUNY Albany

David C. Berliner
Arizona State University
Henry Braun Boston College

Casey Cobb
University of Connecticut

Arnold Danzig
San Jose State University
Linda Darling-Hammond
Stanford University

Elizabeth H. DeBray
University of Georgia

David E. DeMatthews
University of Texas at Austin

Chad d'Entremont Rennie Center
for Education Research & Policy

John Diamond
University of Wisconsin, Madison

Matthew Di Carlo
Albert Shanker Institute

Sherman Dorn
Arizona State University

Michael J. Dumas
University of California, Berkeley

Kathy Escamilla
University of Colorado, Boulder

Yariv Feniger Ben-Gurion
University of the Negev

Melissa Lynn Freeman
Adams State College

Rachael Gabriel
University of Connecticut

Amy Garrett Dikkers University
of North Carolina, Wilmington

Gene V Glass
Arizona State University

Ronald Glass University of
California, Santa Cruz

Jacob P. K. Gross
University of Louisville
Eric M. Haas WestEd

Julian Vasquez Heilig California
State University, Sacramento
Kimberly Kappler Hewitt
University of North Carolina
Greensboro

Aimee Howley Ohio University

Steve Klees University of Maryland
Jaekyung Lee SUNY Buffalo

Jessica Nina Lester
Indiana University

Amanda E. Lewis University of
Illinois, Chicago

Chad R. Lochmiller Indiana
University

Christopher Lubienski Indiana
University

Sarah Lubienski Indiana University

William J. Mathis
University of Colorado, Boulder

Michele S. Moses
University of Colorado, Boulder

Julianne Moss
Deakin University, Australia

Sharon Nichols
University of Texas, San Antonio

Eric Parsons
University of Missouri-Columbia

Amanda U. Potterton
University of Kentucky

Susan L. Robertson
Bristol University

Gloria M. Rodriguez
University of California, Davis

R. Anthony Rolle
University of Houston

A. G. Rud
Washington State University

Patricia Sánchez University of
University of Texas, San Antonio

Janelle Scott University of
California, Berkeley

Jack Schneider University of
Massachusetts Lowell

Noah Sobe Loyola University

Nelly P. Stromquist
University of Maryland

Benjamin Superfine
University of Illinois, Chicago

Adai Tefera
Virginia Commonwealth University

A. Chris Torres
Michigan State University

Tina Trujillo
University of California, Berkeley

Federico R. Waitoller
University of Illinois, Chicago

Larisa Warhol
University of Connecticut

John Weathers University of
Colorado, Colorado Springs

Kevin Welner
University of Colorado, Boulder

Terrence G. Wiley
Center for Applied Linguistics

John Willinsky
Stanford University

Jennifer R. Wolgemuth
University of South Florida

Kyo Yamashiro
Claremont Graduate University

Miri Yemini
Tel Aviv University, Israel

archivos analíticos de políticas educativas
consejo editorial

Editores Asociados: **Felicitas Acosta** (Universidad Nacional de General Sarmiento), **Armando Alcántara Santuario** (Universidad Nacional Autónoma de México), **Ignacio Barrenechea**, **Jason Beech** (Universidad de San Andrés), **Angelica Buendía**, (Metropolitan Autonomous University), **Alejandra Falabella** (Universidad Alberto Hurtado, Chile), **Carmuca Gómez-Bueno** (Universidad de Granada), **Veronica Gottau** (Universidad Torcuato Di Tella), **Carolina Guzmán-Valenzuela** (Universidade de Chile), **Antonia Lozano-Díaz** (University of Almería), **Antonio Luzon**, (Universidad de Granada), **María Teresa Martín Palomo** (University of Almería), **María Fernández Mellizo-Soto** (Universidad Complutense de Madrid), **Tiburcio Moreno** (Autonomous Metropolitan University-Cuajimalpa Unit), **José Luis Ramírez**, (Universidad de Sonora), **Axel Rivas** (Universidad de San Andrés), **César Lorenzo Rodríguez Uribe** (Universidad Marista de Guadalajara), **María Veronica Santelices** (Pontificia Universidad Católica de Chile)

Claudio Almonacid

Universidad Metropolitana de
Ciencias de la Educación, Chile

Miguel Ángel Arias Ortega

Universidad Autónoma de la
Ciudad de México

Xavier Besalú Costa

Universitat de Girona, España

Xavier Bonal Sarro Universidad
Autónoma de Barcelona, España

Antonio Bolívar Boitia

Universidad de Granada, España

José Joaquín Brunner Universidad
Diego Portales, Chile

Damián Canales Sánchez

Instituto Nacional para la
Evaluación de la Educación,
México

Gabriela de la Cruz Flores

Universidad Nacional Autónoma de
México

Marco Antonio Delgado Fuentes

Universidad Iberoamericana,
México

Inés Dussel, DIE-CINVESTAV,
México

Pedro Flores Crespo Universidad
Iberoamericana, México

Ana María García de Fanelli

Centro de Estudios de Estado y
Sociedad (CEDES) CONICET,
Argentina

Juan Carlos González Faraco

Universidad de Huelva, España

María Clemente Linuesa

Universidad de Salamanca, España

Jaume Martínez Bonafé

Universitat de València, España

Alejandro Márquez Jiménez

Instituto de Investigaciones sobre la
Universidad y la Educación,
UNAM, México

María Guadalupe Olivier Tellez,
Universidad Pedagógica Nacional,
México

Miguel Pereyra Universidad de
Granada, España

Mónica Pini Universidad Nacional
de San Martín, Argentina

Omar Orlando Pulido Chaves

Instituto para la Investigación
Educativa y el Desarrollo
Pedagógico (IDEP)

José Ignacio Rivas Flores
Universidad de Málaga, España

Miriam Rodríguez Vargas

Universidad Autónoma de
Tamaulipas, México

José Gregorio Rodríguez

Universidad Nacional de Colombia,
Colombia

Mario Rueda Beltrán Instituto de
Investigaciones sobre la Universidad
y la Educación, UNAM, México

José Luis San Fabián Maroto

Universidad de Oviedo,
España

Jurjo Torres Santomé, Universidad
de la Coruña, España

Yengny Marisol Silva Laya

Universidad Iberoamericana,
México

Ernesto Treviño Ronzón

Universidad Veracruzana, México

Ernesto Treviño Villarreal

Universidad Diego Portales
Santiago, Chile

Antoni Verger Planells

Universidad Autónoma de
Barcelona, España

Catalina Wainerman

Universidad de San Andrés,
Argentina

Juan Carlos Yáñez Velasco

Universidad de Colima, México

arquivos analíticos de políticas educativas
conselho editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editoras Associadas: **Andréa Barbosa Gouveia** (Universidade Federal do Paraná), **Kaizo Iwakami Beltrao**, (Brazilian School of Public and Private Management - EBAPE/FGV), **Sheizi Calheira de Freitas** (Federal University of Bahia), **Maria Margarida Machado**, (Federal University of Goiás / Universidade Federal de Goiás), **Gilberto José Miranda**, (Universidade Federal de Uberlândia, Brazil), **Marcia Pletsch** (Universidade Federal Rural do Rio de Janeiro), **Maria Lúcia Rodrigues Muller** (Universidade Federal de Mato Grosso e Science), **Sandra Regina Sales** (Universidade Federal Rural do Rio de Janeiro)

Almerindo Afonso

Universidade do Minho
Portugal

Alexandre Fernandez Vaz

Universidade Federal de Santa
Catarina, Brasil

José Augusto Pacheco

Universidade do Minho, Portugal

Rosanna Maria Barros Sá

Universidade do Algarve
Portugal

Regina Célia Linhares Hostins

Universidade do Vale do Itajaí,
Brasil

Jane Paiva

Universidade do Estado do Rio de
Janeiro, Brasil

Maria Helena Bonilla

Universidade Federal da Bahia
Brasil

Alfredo Macedo Gomes

Universidade Federal de Pernambuco
Brasil

Paulo Alberto Santos Vieira

Universidade do Estado de Mato
Grosso, Brasil

Rosa Maria Bueno Fischer

Universidade Federal do Rio Grande
do Sul, Brasil

Jefferson Mainardes

Universidade Estadual de Ponta
Grossa, Brasil

Fabiany de Cássia Tavares Silva

Universidade Federal do Mato
Grosso do Sul, Brasil

Alice Casimiro Lopes

Universidade do Estado do Rio de
Janeiro, Brasil

Jader Janer Moreira Lopes

Universidade Federal Fluminense e
Universidade Federal de Juiz de Fora,
Brasil

António Teodoro

Universidade Lusófona
Portugal

Suzana Feldens Schwertner

Centro Universitário Univates
Brasil

Debora Nunes

Universidade Federal do Rio Grande
do Norte, Brasil

Lílian do Valle

Universidade do Estado do Rio de
Janeiro, Brasil

Geovana Mendonça Lunardi

Mendes Universidade do Estado de
Santa Catarina

Alda Junqueira Marin

Pontifícia Universidade Católica de
São Paulo, Brasil

Alfredo Veiga-Neto

Universidade Federal do Rio Grande
do Sul, Brasil

Flávia Miller Naethe Motta

Universidade Federal Rural do Rio de
Janeiro, Brasil

Dalila Andrade Oliveira

Universidade Federal de Minas
Gerais, Brasil