Article

Urban Education
2018, Vol. 53(5) 591–620

DOI: 10.1177/0042085915618710
journals.sagepub.com/home/uex
SAGE

AYP Status, Urbanicity, and Sector: School-to-School Variation in Instruction

Laura Northrop[1] and Sean Kelly[1]

Abstract

This study investigates whether adequate yearly progress (AYP) status, locale, and sector—common variables used to judge the quality of schools—accurately signal true differences in instructional practices in high school mathematics and science. Using data from the High School Longitudinal Study (HSLS), we find the school-to-school variation in instructional practices to be minimal. Controlling for a variety of school and teacher characteristics, we find that there is no difference in the use of developmental instruction between schools that make AYP and schools that do not, urban and nonurban schools, and public and private schools.

Keywords

No Child Left Behind, programs, school choice, urban education, high school

Recent policy changes, such as the school choice reform movement and No Child Left Behind (NCLB), have led educators, parents, and the general public to view schools as critically variable in their effects on student learning. Although schools do differ somewhat, these policies have likely exacerbated perceptions of school-to-school differences in quality (Kelly &

[1]University of Pittsburgh, PA, USA

Corresponding Author:
Laura Northrop, Doctoral Candidate , University of Pittsburgh, 5500 Wesley W. Posvar Hall, 230 S. Bouquet St., Pittsburgh, PA 15260, USA.
Email: lsn6@pitt.edu

Majerus, 2011). NCLB, through the adequate yearly progress (AYP) criteria, attempts to quantify and clarify these differences in school quality. Making AYP is a way for a school to signal that they are of high quality—leaving the multitude of schools labeled "failing" every year with the burden of raising achievement, while convincing parents and the public that they are still high-quality schools worthy of sending their children to. However, recent studies show that while absolute levels of student achievement may vary, learning rates during the school year between students of different background characteristics, and their schools, are quite similar (Downey, von Hippel, & Hughes, 2008; McCoach, O'Connell, Reis, & Levitt, 2006). The observed similarity in learning rates across schools strongly suggests that instructional practices in so-called "failing" schools may in reality be quite similar to the instruction seen in the schools that higher achieving students attend. In this study, we examine school-to-school variation in instructional approaches used in high school mathematics and science, and whether or not those instructional approaches are associated with school-level variables.

In examining teacher instructional approaches, we draw from the growing literature that supports the idea that, within schools, teacher quality is an important predictor of student achievement (Konstantopoulos, 2012; Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004), and that different instructional practices have a disparate impact on student achievement across varying grade levels and subjects (Guarino, Hamilton, Lockwood, & Rathbun, 2006; McCaffrey et al., 2001; Von Secker & Lissitz, 1999). Do students who attend different types of schools have widely varying access to instructional practices that promote engagement and achievement growth? Can the observable qualities of a school that are often used to evaluate schools—AYP status, urban versus nonurban, and sector—strongly signal real differences in instructional quality?

First, we review the literature on school-to-school variation in instruction in general, as well as the link between instructional practices and student achievement in high school math and science specifically. Next, we consider the effect of instruction on student engagement and subsequent academic achievement. Then we review what is known about the effects of AYP status, locale, and sector on teacher instructional practices. Last, using the High School Longitudinal Study (HSLS), we present means-as-outcomes regression models investigating school-to-school differences in instruction. We consider three broad patterns of instruction: developmental instruction (Kelly, 2009; Metz, 1978; Pace & Hemmings, 2007), disciplined inquiry (Newmann, Marks, & Gamoran, 1996), and an emphasis on standardized test preparation. Developmental instruction refers to an engagement-focused approach to instruction emphasizing self-directed, interactive classroom tasks where the teacher is responsive to student ideas, relinquishing some authority over the direction of learning in the classroom. Disciplined inquiry consists of instruction featuring three interrelated emphases: connecting to students' prior knowledge, developing an in-depth understanding of topics, and elaborately communicating this understanding. These abstracted instructional typologies parallel others used in research on students' access to high-quality instruction (e.g., Gamoran & Carbonaro, 2002; Kelly, 2009; Kelly & Majerus, 2011; Raudenbush, Rowan, & Cheong, 1993) and are discussed in greater detail in the "Method" section.

School-to-School Variation in Instructional Practices

A growing body of research supports the idea that, within schools, teacher quality is an important component of student achievement. Konstantopoulos (2012), in summarizing the literature on teacher effects, noted that while "there are substantial differences among teachers in their ability to produce achievement gains" (p. 45), the difference in gain scores is not consistently linked to observable teacher characteristics such as education, experience, licensure, or salary. One possibility is that it is the type of instruction that teachers choose to use that primarily contributes to student achievement, rather than basic presage characteristics (Dunkin & Biddle, 1974; Wayne & Youngs, 2003).[1] Aggregating this hypothesis to the school level, it seems likely that even if schools differ considerably in basic observable measures of teacher quality (e.g., Lankford, Loeb, & Wyckoff, 2002), actual instructional differences might be much more modest, or differ in ways that are not easy to predict.

Prior research on school-to-school differences tends to focus on student achievement outcomes rather than more closely examining instructional processes and generally finds that the majority of variation in student achievement lies within, not between, schools (Coleman et al., 1966; Scheerens & Bosker, 1997). This remains the case even when examining sector differences in achievement between public and private schools (Carbonaro, 2006; Carbonaro & Covay, 2010; Lubienski & Lubienski, 2013). Similarly, studies that have investigated school-to-school variation in instructional practice have found that most variation in use of instructional practices lies within, not between, schools (Gamoran & Carbonaro, 2002; Kelly, 2010a; Kelly & Majerus, 2011; Raudenbush et al., 1993). These studies primarily focused on English language arts instruction, and generally found little variation in students' access to different types of instructional practices, with the between-school variance in teacher instructional practices ranging from 5% to 25%. Within schools, it was often differences in track level of courses (e.g., regular vs. honors English) that accounted for the variation in instruction (Gamoran & Carbonaro, 2002; Raudenbush et al., 1993).

One study that examined math and science instruction at the secondary level found that the variation in instructional practices in math and science was much lower than those found in English (Raudenbush et al., 1993). In this study, the focus was on the emphasis of higher order thinking skills in instructional goals. Raudenbush et al. found less than 1% of the variance in math instruction to be between schools, and only 1.4% of the variance to be between schools in science. However, the effect of tracking had a strong effect on the instructional practices of teachers, with teachers emphasizing more higher order skills in the honors classes than either academic or nonacademic classes. Thus, it was not the school the teachers were placed at, but the track of the class they were teaching, that influenced their instructional decisions.

Likewise, Kelly (2010b) investigated differences in the use of developmental instruction in mathematics in public and Catholic schools. This study found that the school-level variance in mathematics instruction was approximately 6% of the total variance across teachers. In general, teachers were not using much developmental instruction; however, Catholic school teachers were less likely to report using developmental instruction than their public school counterparts.

In all, the limited evidence reviewed here suggests school-to-school differences in instructional practice are relatively modest. Yet, in high school mathematics and science, the limited variation that is observed does appear to translate into differences in student learning. McCaffrey et al. (2001) examined the effects of instructional practice on high school math achievement. In this study, two types of instructional practices were defined: a "Reform Practices" scale that measured inquiry-based instruction, and a "Traditional Practices" scale that included teacher-centered activities such as worksheets, textbooks, and multiple-choice tests. Results showed a positive effect of the use of reform practices when teaching integrated math classes but not traditional ones, such as content-specific classes like algebra or geometry.

As in math, choice of instructional approach in high school science is also related to student achievement (Von Secker & Lissitz, 1999). Using data from the High School Effectiveness Survey, part of the National Educational Longitudinal Study, Von Secker and Lissitz found that two instructional practices, emphasizing laboratory inquiry and decreasing teacher-centered instruction, were associated with higher science achievement. Surprisingly, emphasis on critical thinking was not positively associated with higher science achievement.

Developmental Instruction and Quality of Learning Environment

Developmental instruction is a student-centered approach (Cuban, 1993) to teaching that centers on increasing student engagement with their coursework. In developmental instruction, opportunities for self-direction, student choice, and interaction with peers serve as mechanisms to promote interest and concentration in instructional content (Metz, 1978). Developmental instruction, with its focus on student interest, student choice, and student interactions, influences academic achievement primarily through the mechanism of student engagement.

Elements of developmental instruction that are positively associated with increased student engagement include positive teacher–student relationships, creating a sense of community and belonging, and assigning challenging and interesting tasks (Shernoff, 2013). In turn, increased student engagement is correlated with academic outcomes both directly, through academic engagement that encompasses behaviors such as paying attention, completing homework, and participating in class, and indirectly, through affective engagement that creates a sense of belonging and being valued in the classroom (Finn & Zimmer, 2012). Evidence from a recent meta-analysis supports the idea that positive student–teacher relationships are associated with both increased student engagement and academic achievement in high school (Roorda, Koomen, Spilt, & Oort, 2011). In addition to academic achievement, student engagement at the high school level is associated with other important outcomes, including decreased drop-out rates (Archambault, Janosz, Fallu, & Pagani, 2009; Rumberger & Rotermund, 2012) and decreased participation in high-risk activities, such as drug use (Griffiths, Lilles, Furlong, & Sidhwa, 2012).

More recently, empirical data from the Measures of Effective Teaching (MET) project directly link components of developmental instruction to increased student learning. Teachers who received more favorable results on student surveys that measured such concepts as student interest and engagement ("I like the way we learn in this class") as well as being valued as participants ("My teacher wants us to share our thoughts") had students who learned the most (Measures of Effective Teaching Project, 2012a). In regard to math achievement, students whose teachers scored in the top 25% on student evaluations learned the equivalent of 4.6 months more than students whose teachers scored in the bottom 25% on student evaluations (Measures of Effective Teaching Project, 2012a).

Do Observable School Characteristics Signal School Quality?

AYP Status

Accountability systems are designed to increase student achievement, in part by spurring teachers at "failing" schools to change instructional practices. By exerting pressure on the school and teachers, accountability systems seek to change teachers' behaviors by creating incentives for teachers to teach a certain way and cover specific content (Hamilton, 2003; Hamilton, Stecher, & Yuan, 2008; Kelly, 2012). In addition, AYP status for a given school, at any one point in time, is intended to serve as a signal to parents and others of the quality of instruction typical in that school.

In states that adopted accountability systems tied to standardized testing earlier than the federal law required, such as Kentucky, teachers reported substantially changing the time allotted to tested curriculum and their instructional practices in response to the standardized test (Stecher, Barron, Kaganoff, & Goodwin, 1998). In the case of Kentucky, eighth-grade mathematics teachers reported increasing the amount of time spent on statistics, probability, and algebra, as well as increased time on instructional strategies aligned with the assessments, including open response questions and real-world applications, demonstrating mathematical ideas, and making cross-curricular connections. In Kentucky, in the late 1990s, teachers said that the assessments were the "most potent influence on instruction in math" (Stecher et al., 1998, p. 74).

Later studies, aimed specifically at studying the accountability pressure exerted by NCLB through the mechanism of AYP status, also found that the pressures exerted can lead teachers to change their behaviors in the classroom. This pressure can lead to both desirable and undesirable changes in teacher practices (Hamilton, 2003; Hamilton et al., 2007). In examining the math and science instruction of seventh- and eighth-grade teachers, Hamilton et al. found that desirable changes include aligning instruction with standards and changing instructional practices, while undesirable changes included a narrowing of the curriculum, instruction geared toward tested topics, using formats that mimicked the testing format (such as an increased use of multiple-choice tests), and focusing on students near the proficiency cut score.

In middle school math and science, instructional practice change stemming from AYP pressure included more use of reteaching math topics when students had not mastered them and emphasizing individual instruction based on assessment data (Hamilton et al., 2007). A small percentage of teachers, 9% to 12% depending on the state, reported using less extended math investigations or projects. In addition, teachers' use of test-focused instructional practices was higher in schools that felt pressure to make AYP (Hamilton, 2012).

Although AYP pressure may induce change in teacher instructional practices, it may also be the case that an observed relationship between AYP status and instruction primarily reflects differences in school and district inputs, including resources, teacher quality, or the sociodemographic composition of students. In particular, AYP status is often a function of student composition, not teacher instructional practices. Schools with high populations of Black and Latino students, and schools with high poverty rates, are less likely to make AYP (Kim & Sunderman, 2005). Although schools that make AYP and schools that do not make AYP exhibit similar rates of improvement in their mean proficiency levels (Kim & Sunderman, 2005), schools that do not make AYP have considerably lower student achievement at the start. That is, schools that do not make AYP fail to do so primarily because their students enter school with lower rates of academic proficiency, not because of differences in teacher instructional quality. In addition, at the high school level, making AYP also reflects school characteristics (Balfanz, Legters, West, & Weber, 2007). In a comparison of 2,030 low-performing high schools, Balfanz et al. found that low-performing high schools were more likely to make AYP if they were smaller in overall size, had smaller teacher–student ratios, had fewer subgroups for accountability, and were located in the rural South.

Although we do not expect large differences in instructional practice as a result of AYP status, such policy designations do exert real external pressure on educators; thus, we hypothesize that students attending a school that did not make AYP may have a heavier emphasis on standardized test preparation and less access to developmental instruction than students attending schools who did make AYP.

Additional School-Level Signals: Urbanicity and Sector

In addition to AYP status, locale and sector are often used as proxies for school quality. Parents often seek to find good public schools through the mechanism of buying a home in a particular neighborhood (Holme, 2002), with higher home prices for schools that have higher test scores (Black, 1999; Figlio & Lucas, 2004). Moreover, when moving, parents prefer to move to school districts receiving a higher "report card" rating (Figlio & Lucas, 2004). Because urban schools are less likely to make AYP (Balfanz et al., 2007), it is likely that there is a perception that urban schools are lower quality schools than their nonurban, and more highly rated, counterparts.

In addition to public perception based on locale alone, instructional practices in urban schools may be seen as less effective than in other schools. Although culturally responsive classroom management (Milner & Tenore, 2010), which includes teachers understanding their students' interests outside of the classroom and valuing students' input inside of the classroom, suggests that developmental instructional practices are important for urban students, unfortunately there is evidence that suggests that urban schools are more likely to use a scripted curriculum (Milner, 2013). These scripted curriculums limit teachers' instructional practices by removing teacher professional judgment and teacher choice from the classroom. In particular, math instruction in urban schools tends to be implemented in a way that does not allow for rigorous discussion of mathematical concepts (Boston & Wilhelm, 2017). Thus, we hypothesize that students attending an urban school may have less access to developmental instruction and disciplined inquiry than their nonurban counterparts, although we would expect any differences to be small. Prior research on teacher instructional practices in minority, high poverty schools suggests that while teachers do use less interactive discourse and more seatwork, overall instructional differences are minimal (Kelly, 2010a).

Likewise, sector is also used as a proxy for quality, with parents choosing private schools, particularly secondary schools, as a way to position their children for entrance to elite colleges (Weis, Cipollone, & Jenkins, 2014). While students who attend private school are more likely to take more rigorous math courses such as precalculus and calculus in the higher grades (Carbonaro & Covay, 2010), this specific instance of rigor may not be indicative of global differences in instructional environment. Indeed, prior research suggests private school students may be exposed to a more traditional schooling experience than their public school counterparts. Using the Chicago School Study Data, Kelly (2010b) found that Catholic school teachers were less likely to use developmental instruction than public school teachers. Similarly, in elementary school mathematics, teachers in Catholic schools were more likely to use traditional methods such as textbooks and worksheets, than more reform-oriented methods such as using manipulatives, small group work, or discussing solutions and writing about how they solved problems (Lubienski & Lubienski, 2013). Likewise, there are also differences at the high school level; students taking more traditional, subject-specific math courses such as Algebra 1 or geometry are more likely to have teacher-centered, traditional instructional practices than students in integrated math courses (McCaffrey et al., 2001). Building on Kelly's (2010b) findings from Chicago, it is possible that the HSLS students attending private schools have less access to developmental instruction than their public school counterparts. However, an alternate explanation is that in the more recent HSLS data, because teachers in private schools are not constrained by standardized tests or state-specified curriculum, they may have more freedom to follow student interest when planning their lessons.

In all, based on our review of the prior literature, we expect to find that differences between schools in teachers' instructional practices are minimal compared with the much larger variation within schools. In addition, we hypothesize that observable characteristics, such as AYP status, locale, and sector will be only weakly related to measures of instructional quality, if at all.

Data and Method

In this article, we use the restricted-access HSLS base year data from 2009, a large, nationally representative study which included approximately 25,000 students in Grade 9 enrolled in 950 schools.[2] This study includes surveys of students, teachers, parents, school administrators, and school counselors, with a focus on math and science. In the base year, both math and science teachers were surveyed about their classroom practices, as well as their teacher beliefs and their opinion of their working conditions.

Our analytic sample includes all 950 schools, of which 770 are public schools and 180 are private schools. Instructional variables reported at the student and teacher level were aggregated to the school level; the school administrator questionnaire provided detailed information about the demographics and conditions of the school, including AYP status. Weights at both the student and school level (Ingels et al., 2011) make the results nationally representative of ninth-grade students in 2009 to 2010.[3]

Dependent Variables

To examine school-to-school differences in instructional practices, three types of instructional practices were identified from student and teacher questionnaires: developmental instruction, disciplined inquiry, and standardized test preparation. In so doing, we map these theoretical constructs onto existing variables in a secondary analysis, and although we chose three specific constructs as instructional measures, the variables pertain to general instructional concepts found in many conceptual frameworks. We report basic descriptive differences in each of these instructional constructs across schools. However, in the multivariate analysis of school-to-school differences, due to measurement concerns with the items used to identify disciplined inquiry and standardized test preparation, we focus solely on the developmental instruction construct.

The developmental instruction items used here come from the student surveys, and focus on the incorporation of student ideas into instruction and efforts to capture student interest: "your teacher values and listens to students' ideas" and "your teacher makes math [science] interesting." The alpha reliability for the math developmental instruction measure is .719; for science .716.

In disciplined inquiry, teachers focus on connecting to prior knowledge, developing in-depth understanding of topics, and sharing conclusions through elaborated communication (Newmann et al., 1996). Measures of disciplined inquiry come from the teacher surveys, and focused on measuring the in-depth understanding and elaborated communication aspect of disciplined inquiry. The math construct included five statements, including focusing on reasoning mathematically, connecting mathematics ideas to one another, and explaining ideas effectively (α reliability = .787). The science construct included three statements, including evaluating arguments on scientific evidence, communicating science ideas effectively, and understanding the relationship between science, technology, and society (α reliability = .661).

Standardized test preparation refers to increased use of test preparation tactics, such as reallocating instructional time to tested content, redesigning classroom materials to mimic the format of the high-stakes test, and teaching test-taking strategies (Hamilton et al., 2007; Hamilton et al., 2008). Measures of standardized test preparation come from the teacher surveys. Teachers were asked directly how much emphasis they placed on preparing students for standardized tests. Although prior research shows the accountability movement has increased use of test preparation in classrooms (Hamilton et al., 2007; Hamilton et al., 2008), we found a small positive correlation between the use of test preparation and both developmental instruction (r = .120) and disciplined inquiry (r = .180) in math. Likewise, there is a small positive correlation between test preparation and both developmental instruction (r = .070) and disciplined inquiry (r = .129) in science. It may be that effective teachers are able to integrate test preparation into their classrooms without sacrificing the time spent using more student-centered, engagement-driven instructional practices.

Table 1 presents the mean and standard deviation of the three instructional scales, and illustrates some limitations of the disciplined inquiry and standardized test preparation measures. First, both the disciplined inquiry and standardized test preparation variables had means that were highly skewed toward the maximum value of the measurement scale, and thus also had limited variability (relative to the mean). We do not report the coefficient of variation

Table 1. Teacher Instructional Practice Measures.

Instructional practice	α	*M*[a] (*SD*)	Items
Math			
Developmental instruction	.71 (2 items)	2.07 (.283)	Your math teacher values and listens to students' ideas. Your math teacher makes math interesting.
Disciplined inquiry	.79 (5 items)	3.20 (.316)	Teaching students to reason mathematically. Teaching students how mathematics ideas connect with one another. Teaching students the logical structure of mathematics. Teaching students about the history and nature of mathematics. Teaching students to explain ideas in mathematics effectively.
Standardized test prep	(1 item)	3.24 (.552)	Preparing students for standardized tests.
Science			
Developmental instruction	.72 (2 items)	1.98 (.302)	Your science teacher values and listens to students' ideas. Your science teacher makes science interesting.
Disciplined inquiry	.67 (3 items)	3.19 (.413)	Teaching students to evaluate arguments based on scientific evidence. Teaching students how to communicate ideas in science effectively. Teaching students about the relationship between science, technology, and society.
Standardized test prep	(1 item)	3.09 (.687)	Preparing students for standardized tests.
Math/science combined			
Developmental instruction	.61 (4 items)	2.03 (.218)	Your math teacher values and listens to students' ideas. Your math teacher makes math interesting. Your science teacher values and listens to students' ideas. Your science teacher makes science interesting.

[a]*M*s and *SD* reported at the school level. Scale range is 1 to 4.

(*SD*/mean) in Table 1, but such calculations illustrate the compression of the responses at the top of the distribution. Although the *SD* of the developmental instruction scale is lower in the absolute sense than the disciplined inquiry scale, in both subjects it is actually larger relative to the mean. We suspect that the disciplined inquiry items, which come from the teacher surveys, suffer from social desirability bias, where teachers know that they are supposed to report, for example, that they "teach students how mathematics ideas connect with one another." Standardized test preparation is measured by a single item, and the high mean value may reflect simply the reality of teaching in an era of test-based accountability; the overwhelming majority of teachers report engaging in moderate to heavy test preparation. With only a single item summarizing the extent of a teacher's test preparation, it is difficult to interpret what the limited variability in these responses might mean.

In contrast, the developmental instruction items are centered in the middle of the response distribution, and illustrate somewhat more variability across teachers. To summarize instructional differences across schools, we created a combined math and science scale capturing the average use of developmental instruction across subjects in the same school. Alpha reliability for the combined scale, which included four statements, is .605.

Independent and Control Variables

In examining school-to-school variations in instructional practice, we examine three primary school-level variables that may be related to teachers' instructional practices, AYP pressure, urbanicity, and sector, while controlling for student and teacher characteristics aggregated to the school level.

Our measure of AYP status comes from the administrator survey, which specifically asked, "Is your school currently identified as in need of improvement due to Adequate Yearly Progress (AYP) requirements?" A dichotomous variable was created, with one representing a school that was in need of improvement at any of the five levels (School Improvement Year 1, School Improvement Year 2, corrective action, restructuring, and implementation of restructuring). Of the 770 public schools in the sample, 410 had met AYP requirements, 250 schools were facing sanctions of some type; and 110 schools had missing data on this variable. Of the schools facing AYP sanction, 80 schools were in Year 1, 80 schools were in Year 2, 50 schools were in Year 3, 20 schools were in Year 4, and 20 schools were in Year 5.

Our measure of urbanicity comes from the HSLS variable X1LOCALE, which divides schools into four categories: city, suburb, town, and rural. We created a dummy variable, with "city" schools being coded a 1 and "suburb," "town," and "rural" schools being coded a 0. In this sense, we are conceptualizing an urban school by their physical location, not by the characteristics of the student body or the challenges, real or perceived, that teachers encounter, such as an increased English language learner (ELL) population, higher rates of student poverty, or a lack of parental involvement and support (Milner, 2012). Our measure of geographic locale is a summary measure of urbanicity that yields a heterogeneous mix of schools, not all of which necessarily have factors evocative of urban schools, such as a high proportion of poor minority students or more challenging student behaviors.

School sector refers to the distinction between public and private schools, and differences in governance, religious affiliation, and program emphasis within those sectors. Schools were divided into mutually exclusive categories: traditional public schools (690 schools); charter schools (20 schools); public schools with a special math and science focus (10 schools); public schools with an unknown focus (50 schools); private schools, including both secular and religious affiliations (170 schools); and private schools with an unknown focus (10 schools). There are no private schools with a special math and science focus in this sample.

Control variables include a number of observable characteristics at both the school and teacher level. School characteristics include achievement constructs, school demographics, and school climate. Achievement and related measures of course taking and school progress include the average math achievement score, the percentage of students taking Advanced Placement courses, and the percentage of students who need to repeat ninth grade. School demographic measures include the racial composition of the student body, the mean socioeconomic status (SES) of the student body, the percentage of students on free and reduced lunch and those living below the poverty line, and the percentage of special-needs students such as special education students and ELLs. In addition, educational expectations are measured by a dichotomous variable capturing whether or not the student (or his or her parents) expected the student to complete at least a 4-year college degree. School climate is measured by a scale variable that captures the principal's perception of the level of violence and feeling of safety at the school (Ingels et al., 2011). While the majority of the observable school characteristics were reported at the school level on the administrator's questionnaire, measures of gender, math score, SES, poverty, and student and parent educational expectations were aggregated to the school level from the student-level data.

Teacher characteristics include gender and race, as well as information about the teachers' certification, education level, and years of experience in teaching at the current school. In addition, there are five scales measuring teacher attitudes and perceptions of working conditions at their school. These scales include Teacher's Perceptions of Expectations of the School's Students, Teacher's Perceptions of Professional Community, Teacher Self-Efficacy, Teacher's Perceptions of Principal Support, and Teacher's Perceptions of Collective Responsibility (Ingels et al., 2011). At the school level, aggregate teacher characteristics were taken from the student-level file, such that the results are representative of students' instructional experiences rather than a simple equally weighted count of teachers.

An additional limitation is the survey-based measurement of instructional practice in the HSLS data. Many of the variables come from teacher reports, and it is possible that due to social desirability bias, teachers overreported their use of disciplined inquiry and underreported their use of standardized test preparation. This concern is supported by the high means and limited variability in the disciplined inquiry and standardized test prep scales (see Table 1). Thus, as stated previously, we excluded those measures from our regression analysis. We have greater confidence in our third measure of instructional practice, the developmental instruction scale, because it relied on student reports of engagement and interest in their math and science classes, and thus is unlikely to suffer from the same level of social desirability bias (this is reflected in the mean in Table 1). In addition, research from the MET project found that student perceptions of teacher effectiveness are linked to student achievement growth (Measures of Effective Teaching Project, 2012a).

To summarize, although the measures of instructional practices are limited in several ways, the HSLS data represent an impressive number and range of schools, with reliable measures of a variety of school-level characteristics that go beyond what is typically available in state and local databases. Thus, in all, these data present an important, but limited, snapshot of school-to-school differences in instruction in the era of test-based accountability.

Method

We first present a descriptive analysis of the school-to-school variation in students' access to the three instructional practices, beginning with an analysis of variance that decomposes the variance of instructional practice into between- and within-school components. Next, we explore the differences in instructional practice using a series of t *t*ests to examine differences by observable school and teacher characteristics at the school level. Last, we use a series of means-as-outcomes regression models (Raudenbush & Bryk, 2002), which describe school-to-school differences in developmental instruction as a function of school and teacher characteristics. These models were calculated using Stata's mixed command (Rabe-Hesketh & Skrondal, 2012), with the outcome variable measured at the student level, and all the predictor and control variables measured at the school level. The equations for the mixed-level model thus become

$$Y_{ij} = \beta_{0j} + r_{ij},$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{ (Mean School and Teacher Characteristics)} + \cdots + \gamma_{0Q} + \mu_{0j},$$

where Y_{ij} is the exposure to the instructional practice for student i in school j, β_{0j} is the true mean level of instructional practice at school j, and r_{ij} is the error term. In the Level 2 equation, γ_{00} is the intercept, γ_{01} through γ_{0Q} are the coefficients for the school and teacher characteristics aggregated to the school level, and μ_{0j} is the error term. A series of four models were run, with the first model including only the main predictor variable (AYP status, urbanicity, or sector). The second model controlled for school characteristics, the third model controlled for teacher characteristics, and the fourth was the full model.

In running our regressions, we used Stata's MI command for multiple imputation with five sets of data to handle missing data (Acock, 2012). We did not impute values for our key independent variables, instead creating a category for unknown school type and unknown AYP status.

Results

Variation in Instructional Practices Between and Within Schools

Table 2 presents a decomposition of variance analysis for the three instructional measures, reporting the proportion of variance that lies between schools (the intraclass correlation coefficient [ICC]). In addition to a point estimate of the ICC, Table 2 also reports the 95% confidence interval (CI) for the ICC, and the reliability of the school-level mean for each instructional measure, which is a function of the ICC and the within-school sample size. For each of these measures, the decomposition of variance refers to the same units of analysis, the student at Level 1 and the school at Level 2. In other words, we treat the teacher reports of disciplined inquiry and standardized test prep measures as evidence of the instruction that individual students have access to.

Between-school variation in students' access to developmental instruction is relatively small. Decomposing the variance shows that the majority of the variance in developmental instruction in math and science lies within schools, while only 7% to 9% occurs between schools (see 95% CI column in Table 2). Examining the developmental instructional practices separately by subject, 7% to 9% of the variance in math developmental instruction occur between schools, while 9% to 12% of the variation in science developmental instruction occur between schools. This suggests that teacher and classroom contexts

Table 2. Decomposition of the Variance in Instructional Practices Between and Within Schools.

Instructional practice	Intraclass correlation	Asymptotic *SE*	95% CI	School-level reliability	Average school sample size
Math					
Developmental instruction	.084	.006	[.072, .095]	.649	20.16
Disciplined inquiry	.331	.014	[.304, .358]	.891	16.52
Standardized test preparation	.455	.015	[.430, .484]	.932	16.43
Science					
Developmental instruction	.110	.007	[.097, .124]	.698	18.58
Disciplined inquiry	.459	.015	[.429, .489]	.930	15.72
Standardized test preparation	.611	.014	[.583, .640]	.961	15.53
Math/science combined					
Developmental instruction	.080	.005	[.070, .091]	.649	21.15

Note. CI = confidence interval.

are the primary influences on the use of developmental instruction in high school math and science courses.

In examining the between- and within-school variance for disciplined inquiry and standardized test preparation, we find much larger between-school variation, with between 30% and 60% of the variance occurring between schools. Moreover, the school means of these variables are generally reliably estimated in these data (.9 or above in most cases). However, recall from Table 1 that across individual reports as a whole, both of these measures have high means and low standard deviations. Thus, the higher ICCs of these measures in Table 2 reflect to a large extent simply less total variability in these constructs (which is in the denominator of the ICC). In addition, although variance was measured at the student level, because the measures of these instructional practices came from teacher surveys, all students who have the same teacher will have the same value, also lowering total variability. In general, we worry that a different measurement approach (e.g., observational studies) might yield a much different ICC estimate than data from teacher reports, and thus, we stress the results for developmental instruction, which come from *student reports*.

School-to-School Differences in Instructional Practices at the School Level

Table 3 presents the results from a series of *t* tests comparing schools on our three key variables: made AYP versus did not make AYP, urban versus non-urban schools, and public school versus private school. Although school and teacher characteristics differ between sector and between AYP schools, we see little difference in the use of teacher instructional practices (with the exception of private school teachers), suggesting that, at the school level, the basic school and teacher characteristics measured here do not have a large association with a teacher's choice of instructional practices.

Examining schools that make AYP in comparison with schools that do not make AYP, we see few differences in the instructional practices used by teachers at the school level. Math teachers in AYP sanctioned schools are less likely to use disciplined inquiry than their counterparts at schools that make AYP. However, as anticipated, there are many compositional differences in the students who attend, and teachers who staff, schools that do and do not make AYP. Schools that do not make AYP are more likely to enroll minority students and students from low SES families and families living below the poverty line. Students are less likely to take AP courses, are more likely to repeat ninth grade, and have lower math achievement. Schools that do not make AYP also have significantly more ELL and special education students. Teachers at schools who do not make AYP are less likely to be certified and more likely to have entered teaching through an alternative program. They also exhibit more negative views than their counterparts in schools that make AYP; teachers in AYP sanctioned schools report a lower sense of professional community, less self-efficacy, lower student expectations, less principal support, and a lower collective responsibility.

Considering the differences between urban and nonurban schools, we see few differences in instructional practices. Teachers in nonurban schools are somewhat more likely to report using standardized test preparation in both math and science than their urban counterparts, although teachers in both locales have high means on these measures and place heavy emphasis on standardized test preparation. Although there are few differences between teachers in urban and nonurban schools, there are significant differences in the observable school characteristics. Students in urban schools are more likely to be minority students and more likely to repeat ninth grade. Although urban schools had more ELL students, they enrolled fewer special education students. However, we also found that students in urban schools in our sample had higher SES, were more likely to take AP courses, and had a slightly higher math achievement score.

Table 3. Descriptive Statistics and *t* Tests by AYP Status, Urbanicity, and Sector.[a]

Variables	AYP: No sanction *n* = 410	AYP: Needs improvement *n* = 250	Urban *n* = 270	Nonurban *n* = 680	Public school *n* = 770	Private school *n* = 180
Instructional practices: Math						
Developmental instruction	2.07 (0.258)	2.08 (0.279)	2.08 (0.290)	2.06 (0.280)	2.07 (0.01)	2.01 (0.325)**
Disciplined inquiry	3.20 (0.295)	3.12 (0.314)**	3.20 (0.302)	3.20 (0.322)	3.17 (0.306)	3.31 (0.337)***
Test preparation	3.31 (0.509)	3.30 (0.501)	3.16 (0.543)	3.27 (0.553)**	3.31 (0.503)	2.88 (0.632)***
Instructional practices: Science						
Developmental instruction	1.99 (0.281)	2.00 (0.265)	1.98 (0.305)	1.99 (0.301)	2.0 (0.282)	1.93 (0.028)**
Disciplined inquiry	3.18 (0.379)	3.16 (0.412)	3.21 (0.426)	3.18 (0.407)	3.17 (0.402)	3.29 (0.449)**
Test preparation	3.19 (0.609)	3.14 (0.624)	3.01 (0.737)	3.12 (0.664)*	3.17 (0.023)	2.61 (0.824)***
Instructional practices: Combined						
Developmental instruction	2.06 (0.193)	2.05 (0.208)	2.03 (0.215)	2.03 (0.219)	2.04 (0.203)	1.97 (0.266)***
School characteristics						
Math score	40.38 (5.59)	37.27 (5.33)***	40.88 (6.57)	39.59 (5.60)**	39.15 (5.79)	43.51 (5.15)***
% male students	0.515	0.514	0.505	0.513	0.517	0.483
% free lunch students	0.347 (0.201)	0.506 (0.225)***	0.357 (0.312)	0.333 (0.231)	0.408 (0.226)	0.062 (0.013)***
SES	0.003 (0.369)	−0.216 (0.323)***	0.109 (0.497)	0.001 (0.408)***	−0.084 (0.369)	0.533 (0.352)***
% student below poverty line	0.159	0.255**	0.185	0.167	0.197	0.063***
% ELL students	0.040 (0.063)	0.082 (0.123)***	0.067 (0.105)	0.043 (0.079)***	0.057 (0.093)	0.024 (0.056)***
% SPED students	0.127 (0.060)	0.148 (0.845)***	0.104 (0.101)	0.119 (0.069)*	0.135 (0.070)	0.032 (0.060)***
% students in AP courses	0.161 (0.147)	0.132 (0.103)**	0.190 (0.157)	0.147 (0.125)***	0.151 (0.127)	0.020 (0.163)***
% students White	0.725 (0.257)	0.532 (0.307)***	0.557 (0.301)	0.701 (0.273)***	0.646 (0.293)	0.718 (0.262)**
% students Hispanic	0.118 (0.184)	0.196 (0.237)***	0.194 (0.234)	0.120 (0.183)***	0.149 (0.208)	0.105 (0.167)**
% students Black	0.114 (0.154)	0.211 (0.237)***	0.176 (0.205)	0.131 (0.184)**	0.156 (0.202)	0.091 (0.124)***
% students Asian	0.032 (0.059)	0.045 (0.081)*	0.061 (0.109)	0.032 (0.053)***	0.036 (0.067)	0.056 (0.097)***
% students American Indian	0.010 (0.033)	0.013 (0.066)	0.01 (0.016)	0.012 (0.065)	0.011 (0.046)	0.014 (0.085)

(continued)

Table 3. (continued)

Variables	AYP: No sanction n = 410	AYP: Needs improvement n = 250	Urban n = 270	Nonurban n = 680	Public school n = 770	Private school n = 180
% students repeat ninth grade	0.046 (0.064)	0.082 (0.117)***	0.059 (0.118)	0.044 (0.063)*	0.059 (0.088)	0.003 (0.021)***
% school math/science focus	0.009	0.008	0.008	0.006	0.01	0
% school single sex	0.005	0.004	0.079	0.024***	0.004	0.184***
School climate	−0.473 (0.863)	−0.963 (0.987)***	−0.283 (1.20)	−0.395 (0.008)	−0.654 (0.939)	0.805 (0.609)***
Student educational expectations	0.736 (0.145)	0.671 (0.154)***	0.774 (0.170)	0.731 (0.159)***	0.714 (0.150)	0.872 (0.157)***
Parent educational expectations	0.773 (0.158)	0.725 (0.163)***	0.810 (0.191)	0.781 (0.158)*	0.759 (0.162)	0.922 (0.127)***
Teacher characteristics						
% White	0.899 (0.181)	0.832 (0.244)***	0.841 (0.225)	0.884 (0.211)**	0.867 (0.217)	0.897 (0.211)
% Black	0.034 (0.114)	0.055 (0.139)*	0.043 (0.117)	0.044 (0.136)	0.048 (0.137)	0.021 (0.093)*
% Hispanic	0.033 (0.104)	0.062 (0.160)**	0.053 (0.146)	0.036 (0.123)	0.043 (0.128)	0.034 (0.140)
% Asian	0.016 (0.057)	0.024 (0.065)	0.038 (0.095)	0.016 (0.058)***	0.021 (0.065)	0.027 (0.095)
% Other race	0.019 (0.073)	0.027 (0.089)	0.024 (0.082)	0.019 (0.075)	0.021 (0.075)	0.021 (0.084)
% male	0.423 (0.250)	0.412 (0.248)	0.417 (0.265)	0.413 (0.271)	0.417 (0.254)	0.404 (0.334)
% regular certification	0.855 (0.190)	0.810 (0.013)**	0.759 (0.276)	0.804 (.242)*	0.835 (0.201)	0.591 (0.350)***
% alternative program	0.224 (0.211)	0.289 (0.260)***	0.259 (0.248)	0.247 (0.248)	0.253 (0.237)	0.243 (0.291)
% master's degree	0.547 (0.290)	0.540 (0.283)	0.552 (0.296)	0.529 (0.306)	0.540 (0.291)	0.513 (0.354)
Years at current school	7.54 (3.93)	7.01 (3.61)	7.79 (4.88)	7.71 (4.35)	7.31 (3.90)	9.64 (6.26)***
Community scale	0.035 (0.584)	−0.089 (0.571)**	0.002 (0.653)	−0.002 (0.603)	−0.023 (0.591)	0.104 (0.723)*
Self-efficacy scale	0.060 (0.514)	−0.041 (0.551)*	0.139 (0.583)	0.075 (0.549)	0.014 (0.541)	0.458 (0.496)***
Expectations scale	0.085 (0.540)	−0.154 (0.627)***	0.043 (0.690)	0.075 (0.572)	−0.019 (0.587)	0.454 (0.551)***
Principal support scale	0.041 (0.663)	−0.139 (0.040)***	−0.001 (0.626)	0.002 (0.673)	−0.042 (0.657)	0.198 (0.640)***
Collective responsibility scale	0.043 (0.581)	−0.242 (0.517)***	0.034 (0.699)	0.013 (0.618)	−0.065 (0.583)	0.405 (0.750)

Note. Descriptive data are unweighted. AYP = adequate yearly progress; SES = socioeconomic status; ELL = English language learner; AP = Advanced Placement courses; SPED = special education

[a]*Ms* and *SD* reported at the school level.

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

Table 4. The Prevalence of Developmental Instruction in Math and Science Across Schools by AYP Status, Urbanicity, and Sector After Controlling for School and Teacher Characteristics: Means-as-Outcomes Regression Models.

Variable	Model 1	Model 2	Model 3	Model 4
AYP: Needs improvement	.016 (.017)	−.007 (.018)	−.002 (.018)	−.010 (.018)
AYP: Unknown status	.024 (.025)	.016 (.024)	.004 (.025)	.007 (.024)
Urban	−.007 (.016)	.002 (.016)	−.009 (.016)	.001 (.016)
Public: Charter	−.066 (.066)	−.061 (.059)	−.051 (.062)	−.046 (.057)
Public: Math/science focus	.043 (.055)	.028 (.054)	.042 (.060)	.039 (.060)
Public: Unknown type	.040 (.031)	.036 (.030)	.027 (.030)	.029 (.030)
Private	−.057 (.021)**	−.026 (.030)	−.040 (.027)	−.023 (.032)
Private: Unknown type	−.190 (.064)**	−.160 (.065)*	−.169 (.061)**	−.161 (.066)*

Note. In a set of multiple regressions, the use of developmental instruction was related to three separate independent variables: AYP status, urbanicity, and sector. For each independent variable of interest, four models were run: Model 1, which included only the predictor variable(s); Model 2 which included predictor variable(s) and school characteristics; Model 3 with predictor variable(s) and teacher characteristics; and Model 4 with both school and teacher characteristics. The values in parathensis are standard errors . AYP = adequate yearly progress.
*$p < .05$. **$p < .01$. ***$p < .001$.

Examining the differences between public and private schools, we see statistically significant differences in all three instructional practices for both math and science. Public school teachers are more likely to use developmental instruction, less likely to use disciplined inquiry, and more likely to use standardized test prep than their private school counterparts. There are also differences between sectors in both student and teacher characteristics. Public schools are more likely to enroll minority students and students from lower SES families than private schools. Public school students are also less likely to take AP courses, more likely to repeat ninth grade, and have a lower math achievement. Public schools enroll more than 4 times as many special education students than private schools. Between sectors, teachers varied considerably in their attitudes in teaching; public school teachers reported less professional community, less self-efficacy, lower expectations of students, less principal support, and less collective responsibility than did private school teachers.

Means-as-Outcomes Models of Developmental Instruction at the School Level

Last, we examine a series of means-as-outcomes models to explore the school-to-school variation in developmental instruction in a multivariate framework. Table 4 shows the coefficients for each of the predictor variables across a set of four models: Model 1 includes only the predictor variables, Model 2 controls school characteristics, Model 3 controls teacher characteristics, and Model 4 includes both school and teacher characteristics. Table 5 shows the control coefficients for each of the three predictor variables from the full model, Model 4.

Results from the models indicate that there is no difference in the use of developmental instruction between schools that make AYP and schools that do not, urban and nonurban schools, and public and private schools.[4] Similar to prior research that found private schools tend to use more traditional teaching methods (Kelly, 2010b; Lubienski & Lubienski, 2013), private schools are initially associated with less use of developmental instruction. However, when school and teacher characteristics are included, sector is no longer statistically significant. Although there is still a negative statistically significant association for the 10 private schools of unknown affiliation in the full model, this group represents such a small percentage of schools in the sample (6% of private schools and 1% of all schools) that the number of students effected by this variation is quite limited.

Likewise, there are no associations with either aggregated school characteristics or aggregated teacher characteristics on use of developmental instruction. Interestingly, despite the statistically significant differences in teacher perceptions of school conditions in Table 3 between AYP sanctioned and nonsanctioned schools, and between public and private schools, teacher perceptions were not associated with use of instructional practice. Although teachers at AYP sanctioned schools tend to rate their school climate more negatively, and tend to report having less principal support, and less community with their fellow teachers, these perceptions of working conditions do not seem to be related to the use of instructional practices at the school level.

Given that many urban schools exhibit what are considered to be "urban characteristics," and that urban high schools are more likely to not make AYP (Balfanz et al., 2007), additional analysis (not reported here) examined the interaction effect between urbanicity and AYP on instructional practice. In addition, we ran an analysis investigating the effect of AYP using the subsample of only urban public schools (n = 187), as well as an analysis of all public schools with both the AYP and urbanicity variable in the same model. Similar to the results reported in Table 4, there was a null effect for both the interaction term and AYP status in urban schools on use of developmental instruction.

In addition, we ran our means-as-outcomes models on aggregated combined math/science measures of disciplined inquiry and test preparation (results not shown) and found very few school-to-school differences. Similar to the results for developmental instruction, there are no effects of AYP

Table 5. Coefficients for Control Variables in Model 4: Means-as-Outcomes Models.

Variable	AYP Status	Urbanicity	Sector
School characteristics			
Mean math score	−.005 (.002)*	−.005 (.002)*	−.005 (.002)*
% male students	−.081 (.073)	−.032 (.052)	−.033 (.052)
% free lunch students	.001 (.071)	.050 (.062)	.049 (.064)
SES	.018 (.048)	−.003 (.042)	.014 (.044)
% below poverty	−.118 (.081)	−.167 (.077)*	−.156 (.079)
% ELL students	−.031 (.109)	.000 (.107)	.000 (.106)
% SPED students	−.097 (.137)	−.112 (.117)	−.137 (.118)
% students in AP courses	−.037 (.072)	−.047 (.069)	−.051 (.069)
% Hispanic students	−.018 (.057)	−.068 (.054)	−.063 (.054)
% Black students	.030 (.065)	.013 (.060)	.018 (.061)
% Asian students	.011 (.127)	−.060 (.112)	−.065 (.111)
% American Indian students	.024 (.209)	.094 (.164)	.096 (.162)
% students repeat ninth grade	.149 (.108)	.171 (.113)	.167 (.113)
School climate	−.008 (.010)	−.012 (.009)	−.011 (.009)
Student expectations	−.018 (.068)	−.003 (.068)	.006 (.069)
Parent expectations	−.040 (.067)	−.020 (.065)	−.019 (.065)
Teacher characteristics			
% Black	.039 (.081)	.015 (.077)	.004 (.076)
% Hispanic	.023 (.070)	.016 (.065)	.025 (.064)
% Asian	.224 (.137)	.177 (.118)	.172 (.118)
% Other race	.108 (.151)	.066 (.121)	.068 (.122)
% male	−.060 (.035)	−.055 (.031)	−.055 (.031)
% regular certification	−.059 (.041)	.006 (.033)	−.003 (.034)
% alternative program	.034 (.038)	.035 (.035)	.033 (.034)
% master's degree	−.018 (.029)	−.004 (.027)	−.006 (.027)
Years at current school	.002 (.002)	.001 (.002)	.002 (.002)
Community scale	.025 (.017)	.020 (.015)	.017 (.015)
Self-efficacy scale	−.015 (.016)	−.017 (.016)	−.016 (.016)
Expectations scale	−.019 (.017)	−.022 (.016)	−.019 (.016)
Principal support scale	−.010 (.014)	−.008 (.013)	−.007 (.013)
Collective responsibility scale	.003 (.019)	.002 (.017)	.001 (.017)
Constant	2.42 (.123)***	2.27 (.103)***	2.30 (.108)***
Observations	16,160	19,970	19,970
Groups	770	950	950
Adjusted R^2	.01	.04	.03

Note. Coefficients reported here are from the full model regressions, which correspond to Model 4 in Table 4. The values in parathensis are standard errors . AYP = adequate yearly progress; SES = socioeconomic status; ELL = English language learner; AP = Advanced Placement courses; SPED = special education

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

status, locale, or sector on the use of disciplined inquiry. Attending a private school was negatively associated with test preparation; however, there were no effects of AYP status or locale on the use of test preparation.

Discussion

Contrary to the popular belief that school-to-school differences in instructional quality are both considerable and common, our findings indicate that few of the observable school characteristics that are routinely used to signify school quality—AYP status, locale, and sector—are significantly associated with teachers' use of instructional practices. Examining the prevalence of developmental instruction, the most robust measure of practice considered here, the greatest variation in use of instructional practice continues to be within schools, not between schools. This is consistent with prior research that shows only modest school-to-school variation in instructional practices (Gamoran & Carbonaro, 2002; Kelly, 2010a; Kelly & Majerus, 2011; Raudenbush et al., 1993). Likewise, the aggregate student characteristics of a school exert only a weak influence on the use of developmental instruction. Aggregate teacher characteristics produce similarly null results.

Our study suggests that there is not a tight connection between school compositional variables and the instructional practices teachers use. The dynamic nested-layers model of school organization and student learning (Gamoran, Secada, & Marrett, 2000) is useful in interpreting the results of this study. As in the basic nested-layers model of instruction (Barr & Dreeben, 1983; Bidwell & Kasarda, 1980), student learning is expected to respond to instructional practices. However, the dynamic model shares with the loose-coupling perspective the notion that given teachers' organizational autonomy, a close correspondence between various school resources (including material as well as social resources) and teaching practices cannot be assumed (Meyer & Rowan, 1978; Weick, 1976). Therefore, resources will only influence student learning if they are applied at the classroom level and directly enhance instruction. The dynamic nested-layers model suggests that basic observable factors, such as locale and sector, and AYP status (which is both an enactment of accountability policy and a reflection of school composition) will only be salient to the extent that they directly support teaching practices or indirectly support the resources associated with specific instructional practices. In the results presented here, we do not find evidence of consistent effects of such school-level factors on instruction.

Within the framework of the dynamic nested-layers model, it is still possible that individual, teacher-level variables might produce variation in instruction at the school level due to uneven allocation of teachers across schools. However, we do not see evidence of such teacher sorting effects in these data. One explanation for small school-to-school differences in instruction then is that the majority of teachers are essentially competent in the skills currently demanded of them, even those who teach in "failing" schools. Stated differently, teachers may be, for the most part, similarly trained and prepared in the basic competencies of teaching. As a result, the null effects of AYP status seen here may perhaps reflect that teachers were already engaged in the practices promoted by school accountability. Thus, it is important to acknowledge that examining a more specific teacher skills set may reveal greater disparities across schools. For example, urban schools are more likely to have teachers with less subject-matter mastery (Howard & Milner, 2014), such as lower levels of mathematical knowledge for teaching (Hill & Lubienski, 2007). In addition, there may be important variation between schools in culturally responsive teaching practices (see Gay, 2014) that support diverse students' learning, which we did not examine here.

In light of the overall findings in this analysis, we argue that urban schools have been unjustly negatively affected by the public's perception that failing to make AYP signals weak instructional quality. The public pays attention to school report cards that grade the quality of schools, with consequences for where parents decide to send their children to school (Figlio & Lucas, 2004; Jacobsen, Saultz, & Snyder, 2013). Because urban schools are less likely to make AYP (Balfanz et al., 2007), parents may come to believe that all urban schools are failing or have teachers who use less effective instructional practices. As evidence of this perception, consider that the term "urban school" is often used not to describe the locale of a school, but to describe the racial, socioeconomic, and achievement levels of the students attending the school (Milner, 2012). Consequences of this perception may include parents moving out of urban districts to locations with higher school report card scores (Figlio & Lucas, 2004). Even within urban districts, parents gravitate toward choosing schools with higher report card ratings (Yettick, 2016).

To combat these perceptions, urban schools might need to focus less on their AYP status and school report card scores, and more on conveying what they do instructionally. Spotlighting innovative curriculum and instructional practices may reduce the stigma of lower rated schools. For example, in the Cleveland Municipal School District, MC2STEM is a district-run (noncharter) school that focuses on an engaging science and math curriculum. Although MC2STEM has a rating of effective (as opposed to the higher rankings of excellent or excellent with distinction), their innovative partnership with the local science museum, a technology company, and a university provides a rigorous curriculum with a heavy emphasis on hands-on and applied instruction in laboratory science (Cleveland Municipal School District [CMSD], 2015). This is just one example of how solely using the report card rating as an indicator of the quality of teaching might miss the rich learning opportunities provided by a school.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. More recently, the Measures of Effective Teaching (MET) project examined the connection between teacher instruction and student achievement in math and English language arts in Grades 4 to 8. Although the project used classroom observation rubrics (e.g., Charlotte Danielson's The Framework for Teaching [FFT]) instead of individual instructional practices, all five observational tools used in the project were positively correlated with student achievement. For example, teachers who scored in the bottom quartile on the FFT had students who fell 1 month behind in math, whereas teachers who scored in the top quartile on the FFT had students who were 1.5 months ahead in math (Measures of Effective Teaching Project, 2012b).
2. Observations are rounded to the nearest 10, in accordance with National Center for Education Statistics (NCES) rules regarding restricted-use data.
3. Following Stata's procedures for applying weights using the mixed command (StataCorp, 2013), weights were added at both the student level and the school level. A comparison between four different models (unweighted ordinary least squares [OLS] models at the school level, weighted OLS models at the school level, unweighted multilevel regressions, and weighted multilevel regressions) shows similar results for key variables.
4. Additional models run with the duration of adequate yearly progress (AYP) sanction in Years 1 to 5 as the predictor variable, instead of the dichotomous made AYP/did not make AYP, also had null results.

References

Acock, A. C. (2012). *A gentle introduction to Stata* (3rd rev. ed.). College Station, TX: Stata Press.

Archambault, I., Janosz, M., Fallu, J.-S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence*, *32*, 651-670.

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing high schools? *American Educational Research Journal*, *44*, 559-593.

Barr, R., & Dreeben, R. (1983). *How schools work*. Chicago, IL: University of Chicago Press.

Bidwell, C. E., & Kasarda, J. D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education*, *88*, 401-430.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*, 577-599.

Boston, M. D., & Wilhelm, A. G. (2017). Middle school mathematics instruction in instructionally focused urban districts. *Urban Education*, *52*, 829-861.

Carbonaro, W. (2006). Public-private differences in achievement among kindergarten students: Differences in learning opportunities and student outcomes. *American Journal of Education*, *113*(1), 31-65.

Carbonaro, W., & Covay, E. (2010). School sector and student achievement in the era of standards based reform. *Sociology of Education*, *82*, 160-182.

Cleveland Municipal School District. (2015). *MC2STEM about our school*. Retrieved from http://www.clevelandmetroschools.org//site/Default.aspx?PageID=1264

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC.: U.S. Government Printing Office.

Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890-1990*. New York, NY: Teachers College Press.

Downey, D. B., von Hippel, P. T., & Hughes, M. (2008). Are "failing" schools really failing? Using seasonal comparison to evaluate school effectiveness. *Sociology of Education*, *81*, 242-270.

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. New York, NY: Holt, Rinehart and Winston.

Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *The American Economic Review*, *94*, 591-604.

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97-131). New York, NY: Springer.

Gamoran, A., & Carbonaro, W. J. (2002). High school English: A national portrait. *The High School Journal*, *86*(2), 1-13.

Gamoran, A., Secada, W. G., & Marrett, C. B. (2000). The organizational context of teaching and learning: Changing theoretical perspectives. In M. T. Hallinan (Ed.), *Handbook of the sociology of education* (pp. 37-63). New York, NY: Kluwer Academic.

Gay, G. (2014). Culturally responsive teaching principles, practices, and effects. In H. R. Milner & K. Lomotey (Eds.), *Handbook of urban education* (pp. 353-372). New York, NY: Routledge.
Griffiths, A.-J., Lilles, E., Furlong, M. J., & Sidhwa, J. (2012). The relations of adolescent student engagement with troubling and high-risk behaviors. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 563-584). New York, NY: Springer.
Guarino, C. M., Hamilton, L. S., Lockwood, J. R., & Rathbun, A. H. (2006). *Teacher qualifications, instructional practices, and reading and mathematics gains of kindergartners* (NCES 2006-031). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
Hamilton, L. S. (2003). Assessment as a policy tool. *Review of Research in Education*, *27*, 25-68.
Hamilton, L. S. (2012). Measuring teaching quality using student achievement tests: Lessons from educator's responses to No Child Left Behind. In S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement* (pp. 49-75). New York, NY: Teachers College Press.
Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J. L., . . . Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: Rand.
Hamilton, L. S., Stecher, B. M., & Yuan, K. (2008). *Standards-based reform in the United States: History, research, and future directions*. Santa Monica, CA: Rand.
Hill, H. C., & Lubienski, S. T. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. *Educational Policy*, *21*, 747-768.
Holme, J. J. (2002). Buying homes, buying schools: School choice and the social construction of school quality. *Harvard Educational Review*, *72*(2), 177-205.
Howard, T. C., & Milner, H. R. (2014). Teacher preparation for urban schools. In H. R. Milner & K. Lomotey (Eds.), *Handbook of urban education* (pp. 353-372). New York, NY: Routledge.
Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., . . . Leinwand, S. (2011). *High-School Longitudinal Study of 2009 (HSLS:09). Base-year data file documentation* (NCES 2011-328). Washington, DC: National Center for Education Statistics.
Jacobsen, R., Saultz, A., & Snyder, J. W. (2013). When accountability strategies collide: Do policy changes that raise accountability standards also erode public satisfaction? *Educational Policy*, *27*, 360-389.
Kelly, S. (2009). The Black-White gap in mathematics course taking. *Sociology of Education*, *82*, 47-69.
Kelly, S. (2010a). A crisis of authority in predominantly black schools? *Teachers College Record*, *112*, 1247-1274.
Kelly, S. (2010b). The prevalence of developmental instruction in public and Catholic schools. *Teachers College Record*, *112*, 2405-2440.

Kelly, S. (2012). Understanding teacher effects: Market versus process models of educational improvement. In S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement* (pp. 7-32.) New York, NY: Teachers College Press.

Kelly, S., & Majerus, R. (2011). School-to-school variation in disciplined inquiry. *Urban Education*, *46*, 1553-1583.

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, *34*(8), 3-13.

Konstantopoulos, S. (2012). Teacher effects: Past, present and future. In S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement* (pp. 33-48). New York, NY: Teachers College Press.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*, 37-62.

Lubienski, C. A., & Lubienski, S. T. (2013). *The public school advantage: Why public schools outperform private schools*. Chicago, IL: The University of Chicago Press.

McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education*, *32*, 493-517.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*, 14-28.

Measures of Effective Teaching Project. (2012a). *Asking students about teaching: Student perception surveys and their implementation*. Bill and Melinda Gates Foundation, Measures of Effective Teaching Project. Retrieved from http://met-project.org/downloads/Asking_Students_Practitioner_Brief.pdf

Measures of Effective Teaching Project. (2012b). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill and Melinda Gates Foundation, Measures of Effective Teaching Project. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Metz, M. H. (1978). *Classrooms and corridors: The crisis of authority in desegregated secondary schools*. Berkeley: University of California Press.

Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. W. Meyer (Ed.), *Environments and organizations* (pp. 78-110). San Francisco, CA: Jossey-Bass.

Milner, H. R. (2012). But what is urban education? *Urban Education*, *47*, 556-561.

Milner, H. R. (2013). Scripted and narrowed curriculum reform in urban schools. *Urban Education*, *48*, 163-170.

Milner, H. R., & Tenore, F. B. (2010). Classroom management in diverse classrooms. *Urban Education*, *45*, 560-603.

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education*, *104*, 280-312.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237-257.

Pace, J. L., & Hemmings, A. (2007). Understanding authority in classrooms: A review of theory, ideology and research. *Review of Educational Research*, *77*, 4-27.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (Vol. 1). College Station, TX: Stata Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal*, *30*, 523-553.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*, 247-252.

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*, 493-529.

Rumberger, R. W., & Rotermund, S. (2012). The relationship between engagement and high school dropout. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 491-513). New York, NY: Springer.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. New York, NY: Elsevier Science.

Shernoff, D. J. (2013). *Optimal learning environments to promote student engagement*. New York, NY: Springer.

StataCorp. (2013). *Stata multilevel mixed-effects reference manual*. Retrieved from. http://www.stata.com/manuals13/me.pdf

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 Tech. Rep. No. 482) Los Angeles: University of California, Center for Research on Evaluation, Standards and Student Testing.

Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching*, *36*, 1110-1126.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89-122.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, *21*, 1-19.

Weis, L., Cipollone, K., & Jenkins, H. (2014). *Class warfare: Class, race, and college admissions in top-tier secondary schools*. Chicago, IL: The University of Chicago Press.

Yettick, H. (2016). Information is bliss: Information use by school choice participants in Denver. *Urban Education*, *51*, 859-890.

Author Biographies

Laura Northrop is a doctoral candidate in the Department of Administrative and Policy Studies at the University of Pittsburgh and a lecturer in education at John Carroll University. Her research interests include education policy, middle school literacy instruction, and language arts teacher education.

Sean Kelly is an associate professor in the Department of Administrative and Policy Studies at the University of Pittsburgh. His research focuses on the social organization of schools, the school and classroom environments that promote student engagement, and teacher quality.