Journal of International Education and Leadership
http://www.jielusa.org/
Volume 9 Issue 2 Fall 2019
ISSN: 2161-7252

The Effects of Organizing Teaching by Time, Student Grouping, And Professional Staffing: A National Study of Student Outcomes By Urban, Suburban and Rural Schools

Sue Poppink
Western Michigan University

Xin Ma
University of Kentucky

Jianping Shen
Western Michigan University

For this study, we examined the constructs of time, student groupings, and professional staffing in schools and determined their significance in whether or not schools made Adequate Yearly Progress or were required mandatory improvement. We conducted a logistical regression analysis using a national data set, the School and Staffing Survey (2007-2008). We explored 15 predictor variables and found 10 variables had either positive or negative significance: longer school day, earlier start time, block scheduling, traditional and non-traditional departments, subdivided grades, looping, multi-age grouping, reading coaches, mathematics coaches, and science coaches.

Keywords: Achievement scores by school; adequate yearly progress; mandatory improvement; time in schools; groupings in schools; coaches in schools; National Data Base School and Staffing Survey USA; rural, urban and suburban schools

Educators have long tried to improve school effectiveness through experimenting with organizing teaching in various ways and continue to do so now. We used logistical regression analysis with the national School and Staffing Survey (2007-2008) to examine the relationship between the constructs of (a) time, (b) student and teacher groupings, and (c) professional staffing to determine the signify-cance of whether or not schools made Adequate Yearly Progress and whether or not they were placed on the mandatory improvement list. We analyzed schools by urban, suburban and rural geographic areas. We controlled for school background characteristics including school level, free and reduced-price lunch eligibility rate, and the percentage of minority students. Adequate Yearly Progress (AYP) was a measure required by the No Child Left Behind Act (2002) that held schools and districts accountable for making progress toward all students reaching proficiency on achievement tests. A school was identified in need of mandatory improvement if it failed to make AYP for two consecutive years (Education Week, 2011).

To explore the three constructs, we used 15 predictor variables. To examine the effects of the organization of time, we explored five variables: (a) length of the school day, (b) school days per year, (c) time school day begins, (d) a year-around calendar, and (e)

block scheduling. To examine the effects of student grouping in schools we explore four variables: (a) traditional grades for academic discipline-based schools, (b) grades subdivided into small groups, (c) student groups with the same teacher for two or more years – often referred to as "looping," and (d) multi-age classrooms. To examine the effects of professional instructional support staff we use six variables, which include specialists and coaches organized by the school subject matters of (a) reading, (b) mathematics, and (c) science. Time, student groupings, and professional support simultaneously strongly structure the practice of teaching.

We found many theoretical arguments for the various ways to organize teaching and staffing in schools, but empirical studies that use achievement data are sparse. In addition, other researchers have examined the variables we use, but only in isolation from one another and often with limited samples. We were able to examine multiple variables simultaneously, using nationally representative data collected by the Schools and Staffing Survey (SASS).

Literature Review

We organized our literature review by the constructs of time, groups, and professional staff, and within those, by the 15 variables available in the SASS data.

The Organization of Time in Schools - An entire organization, the National Center on Time & Learning, is dedicated to the study of time because time in schools is of such importance to policy makers and researchers. One way to use time is to increase the amount of time by adding hours to the school day or school year. Other ways time can be manipulated in schools are to change a school's start time; provide shorter, fewer breaks by re-arranging the school calendar, and using block scheduling. In our study, we examined these by looking at the length of the school day, the school days per year, the time the school day started, a year round calendar, and block scheduling.

Extended Time in School: Extended Day and Extended Year - Policy mechanisms that have received intense attention from policy makers are extending the amount of time students spend in school either during the school day, the school year, or both. Woods (2015) found that while the research on time has mixed results concerning expanding learning time, policy-makers at the state, district, and school levels have considered making or have made extensions to the school day or year, and these are increasingly popular policies. The Center for American Progress found that more than 300 initiatives to extend learning time were created between 1991 and 2007 in high-poverty and high-minority schools in 30 states; and, of course, this does not include all initiatives (Patall, Cooper, and Allen, 2010).

As to *a lengthened school day*, Patall et al. (2010) noted that several states, districts, charter school groups, and some individual schools have lengthened the school day. The average school day in the United States is around 6.5 hours per day. After an intensive search, Patall and her colleagues uncovered two studies that suggested a positive impact on student achievement, one by Adelman, Haslam, and Pringle (1996), and another by Wheeler (1987). In reviewing these studies, Patall et al. found that an extended day led to improvement in academic achievement, though the relationship across grade levels, socio-economic status (SES), or for all outcomes were rarely significant.

Two recent studies by Folsom and her colleagues (Folsom et al., 2016; Folsom, Osborne, Osborne-Lamkin, Cooley & Smith, 2017) found growth in Florida's lowest performing schools, which were required to extend the school day, however, this growth did not exceed what would have been expected because of natural variation. In this case, implementation of the extended school day had only been in effect one year, and could change over time. Therefore, though these studies

suggest a positive relationship, the results are not strong.

As to lengthened school year, Patall et al. (2010) found eight studies that examined the relationship between school year and academic achievement; of those, seven showed some evidence of a positive relationship between an extended school year and achievement, though they pointed out that the evidence is relatively weak. They also pointed out that effects may vary by whether or not the students are at-risk of school failure, by subject matter, by ethnicity, and by gender. The studies suggested a small but positive effect for at-risk students, minorities, and female students.

A literature review of evidence from countries in the European Union found that extended time, by either day or year, is a key educational resource, but what matters most is the ways in which allocated time is used (Gromada & Shewbridge, 2016). In agreement, Woods (2015) argued that increases in time need to be accompanied by strong instructional practices. In both cases of lengthened school day and extended school year, Patall and her colleagues (2010) found no evidence to suggest that either policy was a detriment to student achievement. Gromada and Shewbridge (2016), however, found that academic achievement has a curvilinear relationship with time, with diminishing returns.

Time School Day Begins - In the early 1990s, medical researchers determined that teenagers have biologically different sleep-and-wake patterns from other age groups such as preadolescents or adults (citations). Many researchers speculated that a later school start time would increase adolescent achievement. Because school start times are tied to after-school activities and the need for those activities to take place before the dinner hour, few schools pushed back the school start time. The literature on time tends to examine either the school start time, or the sleep duration of students. Our study addresses school start time rather than duration. Keller and associates (2014) conducted one study that addressed school start time and student achievement. They found, contrary to their expectations, that for middle and upper class students, earlier school start times were associated with poorer school performance in elementary schools. The same was not true for students in dis-advantaged schools, as expected.

In a longitudinal study, Wahlstrom (2002) found significant benefits in a later start time, including improved attendance and enrollment rates, less sleeping in class, and less student-reported depression. Though not specifically about achievement test scores, Wahlstrom and her colleagues in Minnesota (Fredrickson, Wahlstrom, & Wrobel, 1999; Wahlstrom, 1999; Wahlstrom & Freeman, 1997) and one study in Rhode Island (Carskadon, Wolfson, Tzischinsky, & Acebo, 1995) found clear statistical evidence that students who do not experience sleep lag syndrome report higher grades, less depression, and fewer at-risk behaviors for dropping out of school (Wahlstrom, 2002).

Year-Round Calendar - Another way practitioners and policy makers have construed time in school is to utilize a year-round calendar. A thorough review by Cooper, Valentine, Charlton, and Melson (2003) indicated that, in the year-round calendar, students attended anywhere from five weeks of school followed by one week off, to 18 weeks of school followed by six weeks off, and several combinations in between. What differentiates the year-round calendar from students who attend summer school is that the year-round calendar is not designed to lengthen the time students spend in schools, but rather designed so students have shorter breaks between times when they are in schools. During the 2000–2001 school year, more than 2.16 million students attended about 3,000 schools that operated without the long summer break (National Association for Year-Round Education, 2000, as cited in Cooper, Valentine, Charlton and Melson, 2003).

The length of the school break is important to practitioners and policymakers for at least

two reasons. One is that those in states with a growing population of students may design schools such that students are in multi-track systems, which can serve more students in the same school buildings by using the buildings continuously throughout the year. Cooper and associates (2003) wrote that 40% of schools using a modified calendar were on multi-track systems.

Another is that many have argued that during a long summer break, students tend to lose what they learned during the previous school year. In a meta-analysis of summer vacation on student achievement, Cooper, Nye, Charlton, Lindsey, and Greathouse (1996) found that summer learning loss equaled at least one month of instruction. However, in general, the year round calendar does not seem to correct for that problem, or at least not clearly. Cooper et al. (2003) found that the evidence concerning the differences between traditional school calendars and modified school calendars and their relationship to student achievement was poor. In addition, the evidence indicated that the modified school calendar increased student achievement was quite weak. They noted that a modified school calendar is associated with positive achievement gains for economically disadvantaged students, indicating that these students may not have as many opportunities for academic learning during the long summer break.

Studies that are more recent have indicated a negative or a very weak relationship between a year round calendar and student achievement. In one study, Graves (2011) found significant negative effects of multi-track year-round calendars on academic achievement for all subgroups, with only the limited English-proficiency student subgroups producing unreliable estimates. A few years later, Graves, McMullen & Rouse (2013) found little support for claims that year-round schooling boosts student achievement. In addition, McMullen and his colleagues (2013 & 2015) analyzed data from Wake County, NC, a county that spread its 180 instructional days evenly across the year. The initial results (2013) suggested that year-round schooling had essentially no impact on academic achievement of the average student. When they examined the data by race, they found no evidence that any racial subgroup benefited from year-round schooling. In 2015, McMullen and colleagues re-analyzed the data using a different regression approach and found evidence of a positive impact of year-round calendars for the lowest-performing students. However, even for these students, the estimated academic impact was small.

Block Scheduling - Block scheduling is a way to modify the traditional schedule of high schools from the traditional six or so hours in one-hour blocks to blocks of 80 or 90 or more minutes. In their review of block scheduling, Zepada and Mayer (2006) explained that the prevalence of schools under block scheduling was high during the 1990s. Though schools may still be experimenting with the length of time of various classes, there are few recent research articles on block scheduling.

Zepada and Mayers' (2006) review included 12 studies that addressed student learning as measured by grade point average or student test scores. The results were decidedly mixed. Some studies found that block scheduling increased student achievement, some found it reduced student achievement, and one found no difference. Zepada and Mayers' review raises questions about the impact of block scheduling on student achievement. More recently, Huelskamp (2014) found that students in a college level introductory biology class performed the same whether or not they had participated in block scheduling biology courses in high school; this was true for males and females.

There is strong interest among policy makers and practitioners regarding how to best use and organize school time. However, empirical studies concerning organizing schools by various time-related dimensions are few and non-conclusive. To this end, our study inquired into the relationship between AYP measurements and the following five time-related factors: (a) length of a school day, (b)

length of school days per year, (c) time a school day begins, (d) a year-round calendar, and (e) block scheduling.

The Organization of Student and Teacher Groupings - Educators often try grouping students and teachers in various ways to improve the academic achievement of students. Tyack (1974) argued that standardization and bureaucratization took place in American schools as they evolved from village to urban schools. Educators began to organize teaching by age-level grouping, and this, among other changes, became the central features of school organization. At the elementary level, this meant the self-contained structure of the classroom, and both the grade- and subject-based structure at the secondary level. Since the middle of the 20th century, teachers and schools have tried multiple ways of grouping students and teachers other than the standardized ways, including changing the way schools are organized and the way classrooms are organized.

Below we review two school level ways of grouping students and two classroom level ways. The school level ways are non-graded elementary schools, and grades divided into groups such as "houses," "families," "teams," or "clusters." The two classroom level ways are looping and multi-age classrooms.

Traditional grades or academic discipline-based departments- In our study, we investigated the relationship of schools' AYP measurements and the traditional way to organize schools, with a variable labelled "traditional grades or academic discipline-based department." If a school answered "yes" to this question, that school was deemed to be organized in a traditional manner. If a school answered "no" to this question, a number of possibilities exist for how a school may be organized. An example of organizing schools in a non-traditional way is through non-graded elementary schools. These schools allow students and teachers to organize their work such that students are placed in a group based on their academic performance, rather than their age. Grouping of students can vary considerably even under the title of "non-graded" schools, which gives some insight as to the variation beyond traditional organizing.

Non-graded elementary schools, promoted by Goodlad and Anderson (1959) as early as the 1950s, have a long history in American education. Many different ways to organize nongraded schools have been entertained and tested. We note here that there is little recent research on the nongraded elementary school, and one reason may be that though it may be practiced, it may go under different configurations and different practices. Gutierrez and Slavin (1992) clarified the variations in the nongraded elementary school by stating, "Sometimes this grouping is done for just one subject, sometimes it is done for many subjects, and sometimes students are placed in self-contained multiage classrooms according to their reading performance or general ability" (p. 334). That is, while teaching by achievement grouping, or flexible grouping, is generally conducted within one age-level classroom, teaching by nongraded grouping is conducted across grade levels and is often conducted across multiple teachers.

Nongraded grouping was tried extensively in the 1950s and 1960s and then enjoyed resurgence in the 1990s (Gutierrez & Slavin, 1992). It comes from a long history in American education of trying to meet the learning needs of all students. As Gutierrez and Slavin explain,

> in their original conception, nongraded elementary schools usually incorporated a curriculum structure called *continuous progress*, in which the skills to be learned in such subjects as reading and mathematics are organized into a hierarchical series of levels covering all the grades involved (usually 1-3 or 1-6). (p. 334)

In reviewing the empirical evidence on various types of groupings, Gutierrez and Slavin (1992) discussed four ways to organize students and the relationship of organization

and student achievement. They found that (a) non-graded programs which involved only one subject showed consistently positive achievement effects; (b) non-graded programs involving multiple subjects or comprehensive programs also showed consistently positive achievement; (c) non-graded programs incorporating individualized instruction or Individually Guided Education (IGE) programs showed difference in student achievement, and (d) complicating non-graded plans with individualized instructions did not improve upon non-graded plans.

We found no recent studies on the non-graded elementary school and this may be because the number of ways to group students is varied and complicated by differentiated instruction, placing students in whole age groups for the beginning of instruction and then in small groups to work on specific skills by achievement, and other complications to the grouping of students.

Grades subdivided into small groups such as "houses," or "families." - In 1989 the Carnegie Council on Adolescent Development released *Turning Points: Preparing American Youth for the 21st Century,* a report that encouraged the use of breaking grades levels into smaller communities to create close, trusting relationships between and among adults and students. One such innovation was to subdivide students into houses, families, teams or clusters. Mansberger (2002) found three studies in which schools organized in smaller units were associated with improved student achievement (Flowers et. al., 2000, 1999; Felner et. al., 1997; Lee & Smith, 1993). As with the non-graded elementary schools, there is not much recent research on this organizational pattern, though these houses can be thought of as schools within schools, and the small schools movement has not been very active in the last decade, when the Gates Foundation was disappointed in the results of such schools (Strauss, 2014).

Looping: Student groups remaining with teachers for two or more years - Looping is a way to organize students and teachers in schools in which the students stay with one teacher or group of teachers for two or more years of schooling. George and Lounsbury (2000) found that student achievement can be boosted through long term teacher-student relationships. Looping can create such a relationship.

Looping can be a subset of multi-grade or multi-age groups, meaning there are few empirical studies specifically on looping. However, at least two studies showed a positive effect. Franz, Thompson, Fuller, Hare, Miller, and Walker (2010) found looping had a positive effect on mathematics achievement for middle school students, particularly for females. They did not find a difference for low SES students. Hill & Jones (2018), using statewide data in North Carolina found small but significant test score gains for looping. The effects were largest for minorities.

Multi-age and multi-grade classrooms - *Multi-age classrooms* contain multiple ages in one classroom and are taught by achievement, through individualistic instruction such as worksheets or centers, or a number of other instructional strategies. *Multi-grade* classrooms, on the other hand, are those in which at least two grades are served, and teaching is geared toward two separate grades, rather than by any other determinations such as achievement. Multi-grade classrooms are often created in response to declining or small enrollment, such that one teacher teaches two grade levels.

Veenman (1995) reviewed the empirical evidence on multi-age and multi-grade groupings extensively. Though the literature has not been reviewed extensively since, his synthesis incorporates decades of reviews which serve in understanding the effects. In his review, Veenman found that neither multi-age nor multi-grade classrooms made much difference in terms of cognitive effects for students. They seem to have neither an advantage nor disadvantage in terms of student

achievement. Veenman used only standardized tests, rather than grades, to determine the cognitive effects, arguing that grades were too subjective.

More recently, two studies that focused on the effects of multi-age groupings on the younger members revealed a positive result. Cascio and Schanzenbach (2016) found that having older classmates increased test scores up to eight years after kindergarten. Kallery and Loupidou (2016) found that younger students' cognitive achievements were increased in such groups.

In our study, based on nationally representative sample, we inquire into the relationship between making AYP and the following four student and teacher grouping methods: (a) an arrangement other than traditional grades or academic discipline-based departments, (b) grades subdivided into small groups, (c) student groups remaining two or more years with the same teacher, or looping, and (d) multi-age classrooms.

Staffing Schools with Instructional Support Staff - The third construct we investigated was the use of professionals to support the work of teaching. The professionals included specialists and coaches in reading, mathematics, and science. Researchers and practitioners often use the terms "specialist" and "coach," while sometimes the term "consultant" is used when describing professional personnel in schools who are neither full time classroom teachers nor traditional administrators. Generally speaking, *specialist* is a term that has long been associated with the Title I program, and has meant professionals who work directly with students, as well as with staff. Specialists are generally associated with the elementary school level (Quatroche, Bean, & Hamilton, 2001). In addition, in some states a "specialist" is a certified professional. In their review of reading specialists, Quatroche and colleagues were not able to find evidence of whether or not specialists contributed to student achievement. On the relationship between specialists and academic achievement, the literature, if it exists, is very thin.

Coach is a term that has been used to mean a professional who does not necessarily have direct responsibility for students, but one who works through and with teachers (a) outside the classroom, (b) both outside and inside the classroom, or (c) inside the classroom only--though this third option is far less likely scenario and is sometimes referred to as "co-teaching" (Shidler, 2009). In the survey we used, coaches work with teachers observing lessons, providing feedback, and demonstrating teaching strategies. Scholars also draw a distinction between "peer coaching" and "content coaching," with "peer coaching" meaning practicing classroom teachers coaching practicing classroom teachers. The more recent definition of "content coaching" is a professional who does not have an assigned classroom, but whose main professional responsibility is to work with teachers. Coaching became more prevalent since the passage of the No Child Left Behind Act which encouraged the use of coaches, particularly through the Reading First program (Koepke, 2017).

In the literature, studies on literacy coaching are more prevalent than mathematics or science coaching. Lockwood, McCombs, and Marsh (2010) defined literacy coaches as "specially trained master teachers who provide leadership for a school's literacy program and offer on-site and ongoing support for teachers so they can improve the literacy skills of students" (p. 373).

Evidence concerning whether or not coaches make a difference in student achievement is sparse. As Lockwood and colleagues determined,

> Although reading coaches are prevalent in many schools across the nation, there is little empirical evidence regarding the nature of the coaching and its effectiveness in changing teacher practice and practically no evidence that links coaching to student achievement. (p. 373)

The definition of coaching has evolved over time, which means that studies over time may have been measuring different activities

under the definition of coaching. A small number of studies on the relationship between the more recent content coaching and student achievement tend to support the newer conception of content coaching as a way to increase student achievement. These studies tend to show a positive effect between coaching and student achievement. Again, most studies were conducted on reading content. Ross (1992) found greater gains in student achievement in classes of teachers who reported more frequent contact with their coaches. Similarly, Marsh, McCombs and Martorell (2010) found significantly higher reading scores related to the frequency with which reading teachers reported that the coach reviewed assessment data with them (either individually or in a group). Lockwood et al. (2010) also found that "For reading, on average across grades, having a state funded coach was associated with statistically significant improvements in average annual gains for the 2003 and 2005 cohorts" (p. 381).

An entire issue of *The Elementary School Journal* (2010) devoted to coaching, and edited by Sailors and Shanklin, contained nine articles, three of which (Bean, Draper, Hall, Vandermolen, & Zigmond, 2010; Biancarosa, Bryk, & Dexter, 2010; Matsumura, Garnier, Correnti, Junker, & Bickel, 2010) concerned student achievement. All three articles found an improvement in student achievement because of coaching. A study by Shidler (2010) also found a positive relationship. Matsumura, Garnier, and Spybrook (2013) found that positive effect of content focused reading coaching had a positive effect on the quality of classroom discussion. However, a carefully constructed study by Garet et al. (2008) found no relationship between coaching and reading student achievement at the end of a treatment year or by the end of the following year. One study on mathematics coaching also found a positive effect (Campbell & Malkus, 2011).

Staffing schools with instructional support personnel tends to be mandated by federal programs such as Title I, and state AYP-related school improvement programs (Koepke, 2017). Current studies on the effect of specialists and coaches are sparse and the studies tend to be small-scale within a particular context. In our study, based on a nationally representative sample, we inquire into the relationship between making AYP and the availability of the following six types of instructional support staff: (a) reading specialist, (b) math specialist, (c) science specialist, (d) reading coach, (b) math coach, and (c) science coach.

Conceptualization and Research Questions - In this study we tested whether some constructs to organize teaching are related to school-level student achievement, with control for school background characteristics including school level, free and reduced-price lunch eligibility rate, and the percentage of minority students. To be specific, we tested whether, and if so, which, elements of time, student and teacher grouping methods, and instructional support staffing, are related to AYP measures. We also investigated these relationships by urban, suburban, and rural schools.

This study has the following unique features. First, we have a rich list of variables that we operationalized under the three constructs of (a) the organization and use of time in schools, (b) the grouping of students with teachers, and (c) staffing patterns in schools with various instructional personnel. Thus, we are able to model the effect of variables of the organization of teaching at the school level simultaneously and, therefore, are able to inquire into the unique effect associated with each variable. In the literature, researchers tend to inquire into the relationship between one variable on school organization, on one hand, and school outcome, on the other. Second, through the process of relative weighting, our sample is nationally representative, providing generalizable findings to the United States. Third, due to the large sample we have, we are able to split the sample into urban, suburban, and rural subsamples to inquire into whether the effect of the organization of teaching could play out differently for various locales. Finally, we conducted the study at the school level with

AYP outcome measures.

Methods

Data

We used the 2007-08 Schools and Staffing Survey (SASS) data for the analysis. The SASS first randomly selected schools and linked each sampled school with its school district. A school is more likely to be sampled if there were a larger number of teachers within the school. The SASS then randomly selected teachers from a list of teachers in each sampled school. Because the original sample of schools was not nationally representative, relative weights were used to generalize our results to the national population of schools. Relative weights have several advantages, including approximating the national population to make the results nationally generalizable, adjusting the sample size to its original one to avoid inflating the test statistics, taking into account a school's selection probability to reduce biases that may result from unit non-response, and making use of available information from external sources to improve the precision of sample estimates (National Center for Education Statistics, 2010).

Among questionnaires for school districts, principals, schools, teachers, and school library media centers, we obtained data from the public school survey that contains information such as grade spans, number of students, staffing patterns, teaching vacancies, high school graduation rates, programs and services, and college application rates (see National Center for Education Statistics, 2010). We classified the 2007-08 SASS national sample of schools into categories of urban, suburban, and rural schools. The resulting sample sizes for our analysis were 1,670 urban schools, 3,220 suburban schools, and 2,570 rural schools (all rounded to the nearest 10 due to reporting requirements for using restricted SASS data). The total sample size was 7,460 public schools.

Variables

Three types of variables were involved in our analysis. Outcome variables measured status in making AYP of each school (see Appendix A). One outcome was a dummy variable of whether a school made AYP in 2006-07 (1 = yes vs. 0 = no). If a school failed to make AYP in two consecutive years or more in the same content area, it was identified as in need of mandatory improvement. The other outcome examined whether a school was identified as such in 2006-07. To make both outcome variables positive in statement style for consistency of interpretation, we coded this dummy outcome as 1 = stayed off the category of schools in need of mandatory improvement (a positive result) vs. 0 = fell into the category of schools in need of mandatory improvement (a negative result).

Control variables functioned to adjust for essential school characteristics in data analysis (see Appendix A). For each school, the proportion of students eligible for free or reduced-price lunch was used as a measure of socioeconomic composition of the school, and the proportion of minority students was used as a measure of racial-ethnic composition of the school. We also adjusted for school type with three dummy variables denoting middle school, high school, and middle and high school combined with elementary school as the reference. School location was not used as a control variable but as a stratification variable to group urban, suburban, and rural schools for independent data analysis.

Predictor variables were the key variables in the present analysis (see Appendix A). The first three variables were continuous, measuring length of a school day (in hours), school days per year (used as the number of measurement units with 10 days as one unit), and time a school day begins (counting minutes from 6 o'clock in the morning and used as the number of measurement units with 10 minutes as one unit). The remaining variables were dichotomous, measuring if a certain instructional practice existed in a school. These variables are self-explanatory. For example, for the dummy variable of traditional grades or academic discipline-based departments, a school that

utilized one of these practices was coded as 1 and a school that did not was coded as 0.

Analyses

Because AYP is a measure of school performance or school behavior, we carried out data analyses at the school level using logistical regression analysis. SASS does not have a student component. Therefore, even though school performance ultimately comes from student performance, our analysis is not a multilevel modeling with students nested within schools. As a result, school was the unit of data analysis for school performance (e.g., making AYP) with the analytical goal to predict outcome variables by predictor variables with adjustment for control variables. Multiple regression analysis is an efficient and effective statistical approach for this purpose. Given that both outcomes were dummy variables, logistic regression analysis was performed. Procedurally, logistical regression analyses were performed on each of the two outcome variables separately for urban, suburban, and rural schools.

Control and predictor variables were entered into a logistic regression model as blocks. According to Cohen and Cohen (1983), this block entry approach is more robust in detecting potential relationships among variables. The first block contained five control variables (a) percent of students eligible for free and reduced-price lunch, (b) percent of minority students, and (c) three dummy variables of school type. The second block contained a total of 15 predictor variables which include (a) length of a school day, (b) school days per year, (c) time a school day begins, (d) a year-round calendar to distribute school days across 12 months, (e) block scheduling, (f) traditional grades or academic discipline-based departments, (g) grades subdivided into small groups, (h) student groups with the same teacher for two or more years, (i) multi-age grouping, (j) reading specialist, (k) math specialist, (l) science specialist, (m) reading coach, (n) math coach, and (o) science coach. By entering the control block first and then estimating the effects of variables in the predictor block over and above the control block, any statistically significant predictor variables are considered robust in their effects.

Results

We examined the effects of predictor variables descriptive of characteristics of schools' organization of teaching among urban, suburban, and rural schools separately. We reported descriptive statistics of outcome, control, and predictor variables across the three school locations in Appendix B. For each location, schools were examined on making AYP and staying off the category of schools in need of mandatory improvement (the two outcome variables).

Urban Schools and the Effects of the Organization of Teaching - Table 1 presents results of logistic regression analysis on the effects of schools organization of teaching on whether schools made AYP among urban schools. Five variables were significant in both absolute and relative effects, including length of a school day, time a school day begins, block scheduling, traditional grades or academic discipline-based departments, block scheduling, and reading coach. This stability indicated that even after adjustment over control variables and even in the presence of other predictor variables, these five predictor variables remained statistically significant. Predictor variables like these can be considered robustly important to the outcome measure.

The five variables had the follow effects.

- A longer school day had positive effects. Specifically, consider two urban schools that differ by one hour in a school day, the school with a longer school day was 1.284 times as likely to make AYP than the school with a shorter school day.
- An earlier start time had positive effects in urban schools. If two urban schools differ by 10 minutes in terms of time a school day begins, the school that started the day earlier was 1.107 times (1 ÷ .903) as likely

to make AYP as the school that started the day later.

- Urban schools without block scheduling were 1.524 times (1 ÷ .656) as likely to make AYP as urban schools with block scheduling.
- Urban schools without traditional grades or academic discipline-based departments were 1.733 times (1 ÷ .577) as likely to make AYP as urban schools with traditional grades or academic discipline-based departments.
- Urban schools without reading coaches were 1.605 times (1 ÷ .623) as likely to make AYP as urban schools with reading coaches.

Table 1

Effects of Organizing Teaching by Time among Urban Schools and Adequate Yearly Progress

Variables	Absolute effects			Relative effects		
	B	SE	Exp	B	SE	Exp
Length of a school day (in hours)	.295*	.089	1.343	.250*	.094	1.284
School days per year (in measurement unit of 10 days)	-.026	.100	.974	-.040	.103	.961
Time a school day begins (in measurement unit of 10 minutes)	-.112*	.022	.894	-.102*	.023	.903
A year-round calendar to distribute school days across 12 months (D)	-.210	.156	.811	-.131	.162	.877
Block scheduling (D)	-.428*	.132	.652	-.421*	.138	.656
Traditional grades or academic discipline-based departments	-.545*	.262	.580	-.549*	.277	.577
Grades subdivided into small groups (D)	-.349*	.163	.706	-.283	.171	.754
Student groups with the same teacher two or more years (D)	.124	.167	1.132	.214	.184	1.239
Multi-age grouping (D)	.072	.180	1.075	.118	.196	1.125
Reading specialist (D)	-.190	.126	.827	-.179	.150	.836
Math specialist (D)	-.025	.141	.975	.139	.193	1.149
Science specialist (D)	.168	.176	1.182	.334	.218	1.396
Reading coach (D)	-.547*	.129	.579	-.474*	.158	.623
Math coach (D)	-.389*	.138	.678	-.145	.189	.865
Science coach (D)	-.282	.205	.754	-.031	.256	.969
R square					.215	
Percent of correct prediction of 0					.376	
Percent of correct prediction of 1					.906	
Overall percent of correct prediction					.727	

- $p < .05$.
- Note. Coding for making AYP is 1 = Yes and 0 = No. B = unstandardized coefficient. SE = standard error. Exp = *e* raised to the power of B. D = dichotomous variables with coding as 1 = Yes and 0 = No.

Grades subdivided into small groups and math coach showed statistically significant absolute effects only. In the presence of other predictor variables, the effects of grades subdivided into small groups and math coach became so secondary as to cease to be statistically significant (without statistically significant relative effects). Predictor variables like these can be considered not important to the outcome measure.

Table 1 contains fitting statistics to evaluate the adequacy of model-data-fit. The logistic regression model explained 21.5% of total variance in the outcome measure of making AYP within urban schools. Given that school organization is only one of many school climatic variables that affect schooling outcomes (see Teddlie & Reynolds, 2000), it is quite adequate from a practical perspective for school organization to explain away more than one fifth of the total variance. Overall, the model correctly predicted whether schools made AYP 72.7 out of 100 predictions, which is also adequate from a statistical perspective. The model was more effective in predicting schools that made AYP (90.6 out of 100 predictions) than schools that did not make AYP (37.6 out of 100 predictions).

Table 2 estimates the effects of school organization on whether urban schools stayed off the category of mandatory improvement among urban schools. Three variables were robustly important to the outcome measure: length of a school day, grades subdivided into small groups, and reading coach demonstrated statistically significant absolute and relative effects.

- Consider two urban schools that differ by one hour in a school day, the school with a longer school day was 1.280 times as likely to stay off the category of mandatory improvement as the school with a shorter school day.
- Urban schools without grades subdivided into small groups were 1.570 times (1 ÷ .637) as likely to stay off the category of mandatory improvement as urban schools with grades subdivided into small groups.
- Finally, urban schools without reading coaches were 1.748 times (1 ÷ .572) as likely to stay off the category of mandatory improvement as urban schools with reading coaches.

Time a school day begins, and the presence of a math specialist or math coach demonstrated statistically significant absolute effects only. Without statistically significant relative effects, these predictor variables were not important to the outcome measure.

In Table 2, the logistic regression model accounted for 22.3% of total variance in the outcome measure of staying off the category of mandatory improvement in urban schools. Overall, the model correctly predicted whether schools stayed off the category of mandatory improvement 76.8 out of 100 predictions. The model was far more effective in predicting schools that stayed off the category of mandatory improvement (93.9 out of 100 predictions) than schools that fell into the category (28.1 out of 100 predictions).

Table 2

Effects to Organizing Teaching by Time among Urban Schools and Staying off the Mandatory Improvement Category

Variables	Absolute effects			Relative effects		
	B	SE	Exp	B	SE	Exp
Length of a school day (in hours)	.229*	.104	1.257	.247*	.107	1.280
School days per year (in measurement unit of 10 days)	-.094	.124	.910	-.107	.123	.898
Time a school day begins (in measurement unit of 10 minutes)	-.052*	.025	.950	-.043	.025	.958
A year-round calendar to distribute school days across 12 months (D)	.015	.184	1.015	.058	.190	1.060
Block scheduling (D)	-.187	.145	.829	-.142	.149	.868
Traditional grades or academic discipline-based departments (D)	-.177	.286	.838	-.153	.294	.858
Grades subdivided into small groups (D)	-.512*	.170	.600	-.451*	.176	.637
Student groups with the same teacher two or more years (D)	-.142	.184	.867	-.063	.197	.939
Multi-age grouping (D)	.047	.177	1.048	.152	.187	1.164
Reading specialist (D)	-.234	.144	.791	-.007	.167	.993
Math specialist (D)	-.363*	.160	.696	-.304	.208	.738
Science specialist (D)	-.099	.211	.906	.154	.257	1.166
Reading coach (D)	-.569*	.147	.566	-.559*	.183	.572
Math coach (D)	-.407*	.155	.666	.033	.218	1.034
Science coach (D)	-.394	.222	.675	-.076	.281	.926
R square					.223	
Percent of correct prediction of 0					.281	
Percent of correct prediction of 1					.939	
Overall percent of correct prediction					.768	

- $p < .05$.
- Note. Coding for making AYP for two or more years is 1 = Yes and 0 = No. B = unstandardized coefficient. SE = standard error. Exp = *e* raised to the power of B. D = dichotomous variables with coding as 1 = Yes and 0 = No

Suburban Schools and Effects of the Organization of Teaching - Table 3 presents results of logistic regression analysis on the effects of school organization on whether schools made AYP among suburban schools. Two variables demonstrated statistically significant absolute and relative effects: reading coach and science coach. These predictor variables were robustly important to the outcome measure.

- Suburban schools without reading coaches were 1.541 times (1 ÷ .649) as likely to make AYP as suburban schools with reading coaches.
- Suburban schools with science coaches were 2.016 times as likely to make AYP as suburban schools without science coaches.

Table 3

Effects of Organizing Teaching by Time among Suburban Schools and Adequate Yearly Progress

Variables	Absolute effects			Relative effects		
	B	SE	Exp	B	SE	Exp
Length of a school day (in hours)	.114	.083	1.121	.094	.085	1.098
School days per year (in measurement unit of 10 days)	.069	.105	1.072	.070	.106	1.072
Time a school day begins (in measurement unit of 10 minutes)	-.003	.017	.997	-.003	.018	.997
A year-round calendar to distribute school days across 12 months (D)	-.198	.132	.821	-.174	.134	.840
Traditional grades for academic discipline-based departments (D)	-.002	.188	.998	-.008	.192	.992
Grades subdivided into small groups (D)	-.238	.129	.788	-.254	.132	.776
Student groups with the same teacher two or more years (D)	.056	.139	1.057	.108	.146	1.114
Multi-age grouping (D)	.054	.142	1.056	.059	.147	1.060
Block scheduling (D)	-.072	.104	.930	-.040	.106	.961
Reading specialist (D)	.015	.096	1.015	.089	.107	1.093
Math specialist (D)	-.005	.122	.995	-.024	.147	.976
Science specialist (D)	.200	.175	1.221	.088	.199	1.092
Reading coach (D)	-.362*	.105	.696	-.432*	.121	.649
Math coach (D)	-.145	.133	.865	-.182	.177	.834
Science coach (D)	.391*	.196	1.479	.701*	.242	2.016
R square					.160	
Percent of correct prediction of 0					.128	
Percent of correct prediction of 1					.977	
Overall percent of correct prediction					.798	

- $p < .05$.
- Note. Coding for making AYP is 1 = Yes and 0 = No. B = unstandardized coefficient. SE = standard error. Exp = *e* raised to the power of B. D = dichotomous variables with coding as 1 = Yes and 0 = No.

Table 3 indicates that the logistic regression model explained 16.0% of total variance in the outcome measure of making AYP in suburban schools. Overall, the model correctly predicted whether suburban schools made AYP 79.8 out of 100 predictions. The model was far more effective in predicting suburban schools that made AYP (97.7 out of 100 predictions) than schools that did not make AYP (12.8 out of 100 predictions).

Table 4 estimates the effects of school organization on whether suburban schools stayed off the category of mandatory improvement among suburban schools. Four predictor variables had significant absolute and relative effects: Time a school day begins, grades subdivided into small groups, multi-age grouping, and reading coach.

These predictor variables were robustly important to the outcome measure:

- Consider two suburban schools that differ by 10 minutes in terms of time a school day begins, the school that started the day later was 1.047 times as likely to stay off the category of mandatory improvement as the school that started the day earlier.
- Suburban schools without grades subdivided into small groups were 1.468 times (1 ÷ .681) as likely to stay off the category of mandatory improvement as suburban schools with grades subdivided into small groups.
- Suburban schools without multi-age grouping were 1.403 times (1 ÷ .713) as likely to stay off the category of mandatory improvement as suburban schools with multi-age grouping.

Suburban schools without reading coaches were 1.742 times (1 ÷ .574) as likely to stay off the category of mandatory improvement as suburban schools with reading coaches. The presence of a math specialist and math coach demonstrated statistically significant absolute effects only. These predictor variables were not important to the outcome measure. Science coach demonstrated statistically significant relative effects only. Predictor variables like this require the presence of other predictor variables to be statistically significant, and thus can be considered conditionally important to the outcome measure.

Table 4 indicates that the logistic regression model accounted for 19.8% of total variance in the outcome measure of staying off the category of mandatory improvement within suburban schools. Overall, the model correctly predicted whether schools stayed off the category of mandatory improvement 87.5 out of 100 predictions. The model was far more effective in predicting suburban schools that stayed off the category of mandatory improvement (99.0 out of 100 predictions) than suburban schools that fell into the category (11.9 out of 100 predictions).

Table 4

Effects of Organizing Teaching by Time among Suburban Schools and Staying off the Mandatory Improvement Category

Variables	Absolute effects			Relative effects		
	B	SE	Exp	B	SE	Exp
Length of a school day (in hours)	.069	.109	1.071	.090	.111	1.095
School days per year (inn measurement unit of 10 minutes)	.007	.121	1.007	-.003	.123	.997
Time a school day begins (in measurement unit of 10 minutes)	.044*	.023	1.045	.046*	.023	1.047
A year-round calendar to distribute school days across 12 months (D)	-.146	.162	.864	-.100	.167	.904
Block scheduling (D)	-.123	.125	.884	-.067	.128	.935
Traditional grades or academic discipline-based departments (D)	.140	.233	1.151	.085	.236	1.088
Grades subdivided into small groups (D)	-.396*	.153	.673	-.384*	.157	.681
Student groups with the same teacher two or more years (D)	-.139	.181	.870	.094	.191	1.099
Multi-age grouping (D)	-.347*	.146	.706	-.339*	.152	.713
Reading specialist (D)	-.036	.124	.965	.183	.137	1.201
Math specialist (D)	-.348*	.151	.706	-.297	.191	.743
Science specialist (D)	-.050	.230	.951	.064	.272	1.066
Reading coach (D)	-.557*	.130	.573	-.555*	.157	.574
Math coach (D)	-.451*	.150	.637	-.399	.207	.671
Science coach (D)	.163	.228	1.177	.856*	.284	2.353
R square					.198	
Percent of correct prediction of 0					.119	
Percent of correct prediction of 1					.990	
Overall percent of correct prediction					.875	

- $p < .05$.
- Note. Coding for making AYP for two or more years is 1 = Yes and 0 = No. B = unstandardized coefficient. SE = standard error. Exp = *e* raised to the power of B. D = dichotomous variables with coding as 1 = Yes and 0 = No.

Rural Schools and the Effects of the Organization of Teaching - Table 5 presents results of logistic regression analysis on the effects of school organization on whether schools made AYP among rural schools. Length of a school day demonstrated statistically significant absolute and relative effects. This predictor variable was robustly important to the outcome measure.

- Specifically, consider two rural schools that differ by one hour in a school day, the school with a longer school day was 1.555 times as likely to make AYP as the school with a shorter school day.

On the other hand, the presence of a reading coach and math coach showed statistically significant absolute effects only. These predictor variables were not important to the outcome measure.

Table 5

Effects of School Organization of Teaching among Rural Schools and Making Adequate Yearly Progress

Variables	Absolute effects			Relative effects		
	B	SE	Exp	B	SE	Exp
Length of a school day (in hours)	.432*	.121	1.540	.442*	.127	1.555
School days per year (in measurement unit of 10 minutes)	-.146	.123	.864	-.131	.133	.877
Time a school day begins (in measurement unit of 10 minutes)	-.015	.034	.985	.014	.035	1.014
A year-round calendar to distribute school days across 12 months (D)	.025	.191	1.025	.034	.199	1.035
Block Scheduling (D)	-.077	.156	.926	-.049	.161	.953
Traditional grades or academic discipline-based departments (D)	.135	.303	1.144	.009	.317	1.009
Grades subdivided into small groups (D)	-.002	.208	.998	.013	.215	1.013
Student groups with the same teacher two or more years (D)	.304	.221	1.356	.407	.234	1.502
Multi-age grouping (D)	-.166	.178	.847	-.266	.190	.767
Reading specialist (D)	-.049	.139	.952	-.055	.158	.947
Math specialist (D)	.105	.180	1.111	.139	.227	1.149
Science specialist (D)	.269	.263	1.309	.226	.305	1.254
Reading coach (D)	-.331*	.160	.718	-.243	.184	.784
Math coach (D)	-.440*	.212	.644	-.222	.270	.801
Science coach (D)	-.500	.310	.607	-.258	.364	.773
R square					.188	
Percent of correct prediction of 0					.148	
Percent of correct prediction of 1					.974	
Overall percent of correct prediction					.825	

- $p < .05$.
- Note. Coding for making AYP is 1 = Yes and 0 = No. B = unstandardized coefficient. SE = standard error. Exp = *e* raised to the power of B. D = dichotomous variables with coding as 1 = Yes and 0 = No.

As shown in Table 5, the logistic regression model explained 18.8% of total variance in the outcome measure of making AYP. Overall, the model correctly predicted whether rural schools made AYP 82.5 out of 100 predictions. The model was far more effective in predicting rural schools that made AYP (97.4 out of 100 predictions) than schools that did not make AYP (14.8 out of 100 predictions).

Table 6 estimates the effects of school organization on whether rural schools stayed off the category of mandatory improvement. Student groups with the same teacher two or more years, multi-age grouping, reading coach, and math coach demonstrated statistically significant absolute and relative effects. These predictor variables were robustly important to the outcome measure.

- Rural schools where student groups remained with the same teacher two or more years were 2.209 times as likely to stay off the category of mandatory improvement as rural schools where student groups did not remain with the same teacher two or more years.
- Rural schools without multi-age grouping were 1.672 times (1 ÷ .598) as likely to stay off the category of mandatory improvement as rural schools with multi-age grouping.
- Rural schools without reading coaches were 1.582 times (1 ÷ .632) as likely to stay off the category of mandatory improvement as rural schools with reading coaches.
- Finally, rural schools without math coaches were 1.681 times (1 ÷ .595) as likely to stay off the category of mandatory improvement as rural schools with math coaches.

The presence of a reading specialist demonstrated statistically significant absolute effects only. This predictor variable was not important to the outcome measure. Length of a school day demonstrated statistically significant relative effects only. This predictor variable was conditionally important to the outcome measure.

As shown in Table 6, the logistic regression model accounted for 22.1% of total variance in the outcome measure of staying off the category of mandatory improvement within rural schools. Overall, the model correctly predicted whether rural schools stayed off the category of mandatory improvement 89.0 out of 100 predictions. The model was far more effective in predicting rural schools that stayed off the category of mandatory improvement (99.0 out of 100 predictions) than schools that fell into the category (13.4 out of 100 predictions).

Table 6

Effects of School Organization of Teaching among Rural Schools on Staying off the Mandatory Improvement Category

Variables	Absolute effects			Relative effects		
	B	SE	Exp	B	SE	Exp
Length of a school day (in hours)	.204	.116	1.226	.237*	.121	1.267
School days per year (inn measurement unit of 10 days)	-.018	.115	.982	.004	.117	1.004
Time a school day begins (in measurement unit of 10 minutes)	.009	.033	1.009	.024	.035	1.025
A year-round calendar to distribute school days across 12 months	-.090	.200	.914	-.022	.207	.979
Block scheduling (D)	-.175	.149	.839	-.136	.154	.873
Traditional grades or academic discipline-based departments (D)	-.181	.315	.834	-.330	.327	.719
Grades subdivided into small groups (D)	-.312	.199	.732	-.317	.206	.728
Student groups with the same teacher two or more years (D)	.523*	.235	1.687	.793*	.247	2.209
Multi-age grouping (D)	-.402*	.161	.669	-.514*	.170	.598
Reading specialist (D)	-.346*	.144	.708	-.282	.161	.754
Math specialist (D)	-.058	.187	.944	.417	.252	1.518
Science specialist (D)	-.263	.248	.769	-.456	.304	.634
Reading coach (D)	-.606*	.155	.546	-.459*	.182	.632
Math coach (D)	-.531*	.192	.588	-.519*	.250	.595
Science coach (D)	.028	.323	1.029	.691	.369	1.996
R square					.221	
Percent of correct prediction of 0					.134	
Percent of correct prediction of 1					.990	
Overall percent of correct prediction					.890	

$p < .05$.

Note. Coding for making AYP for two or more years is 1 = Yes and 0 = No. B = unstandardized coefficient. SE = standard error. Exp = *e* raised to the power of B. D = dichotomous variables with coding as 1 = Yes and 0 = No.

Discussion

We defined the organization of teaching by school locations and by the organizing concepts of time, student and teacher grouping, and the professional staffing. In order to discuss the major results, below we recap the results by location, and then review the findings across locations as well as by the organizing concepts and discuss what the literature has to say about the variables. We then discuss implication for policy and practice.

As an overview, Table 7 shows that of the 15 predictive variables, 10 we found to have either a positive or a negative effect on helping an urban, suburban, or rural school either make Adequate Yearly Progress (AYP) stay clear of

the need for mandatory improvement, or both. Looking across concepts (time, grouping, and professional staff), we find that of the 15 variables, three were significant in terms of time, four for school groupings, and three for professional staff.

Table 7

Positive or Negative Significance by Predictor Variables

	Urban		Suburban		Rural	
	Yes AYP	Avoided Mandatory	Yes AYP	Avoided Mandatory	Yes AYP	Avoided Mandatory
Longer School Day	Pos	Pos			Pos	
School Days per Year						
Earlier Start Time	Pos			Neg		
Year-round Calendar						
Block Scheduling	Neg					
Traditional Grade or Academic Discipline-based Departments	Pos					
Grades Subdivided into Small Groups		Neg		Neg		
Student Groups with Same Teacher Two or More Years						Pos
Multi-Age Grouping				Neg		Neg
Reading Specialist						
Mathematics Specialist						
Science Specialist						
Reading Coach	Neg	Neg	Neg	Neg		Neg
Mathematics Coach						Neg
Science Coach			Pos			

Summary of Major Findings by Location - Consistent across the six equations, variables on school organization explain 16.0% to 22.3% of the variance. The amount of variance explained tends to range from small to medium. Thus, we do not overestimate the effects of school organization; nor do we underestimate them. Below we provide an examination of the significant variables.

For *urban* schools, two variables were associated with both AYP and staying clear of the mandatory improvement category. One variable, a longer school day, helped schools make AYP and to stay clear of requiring mandatory improvement requirements. Another variable, reading coaches, worked against schools making AYP and staying clear of mandatory improvement.

An additional three variables were associated with making AYP only. Two variables, traditional grades or academic discipline-based departments, and an earlier start time helped schools make AYP. One of those variables, block scheduling, were negatively associated with making AYP.

One variable was only associated with staying off the mandatory improvement list: grades subdivided into small groups had a significant negative impact.

For *suburban* schools, five variables were significant. One variable, having a reading coach was negatively associated with both making AYP and staying clear of the mandatory achievement test. One variable, having a science coach, was associated with making AYP only. Two variables were negatively associated with only staying clear of the mandatory improvement category: grades subdivided into small groups, and multi-age grouping. One variable, an earlier start time, was negatively associated with staying off the mandatory achievement test.

For *rural* schools, five variables were significant. A lengthened school day was positively associated with making AYP. Four variables were associated with staying off the mandatory achievement list. Three of the variables, multi-age grouping, having a reading coach, and a math coach were negatively associated with staying off the category of mandatory improvement. One variable, students staying with the same teacher for two or more years, was positively associated staying out of the mandatory improvement category.

Major Findings by Professional Staffing, Time and Groupings - Above, we examined the variables by location. Another way to discuss the impact of predictor variables is to count their frequency of a predictor variable that demonstrated statistical significance across locations. In this method of analysis, a predictor variable could have up to six counts of statistical significance (two outcome measures across three school locations). In addition to examine the variables in three grouped forms: time in schools, student grouping, and professional staffing.

From this perspective, the most striking finding is that the presence of reading coaches demonstrated five scenarios of statistical significance (out of six) of having a negative relationship to the achievement of making Adequate Yearly Progress. Thus, we begin the discussion with staffing.

Professional staffing - It is puzzling to understand how it is that this analysis shows reading coaches have a negative effect on a school's ability to make AYP and staying off of the mandatory improvement category in both urban and suburban schools, and to stay off the mandatory improvement for rural schools.

Coaches work with teachers, and coaching includes observing lessons, providing feedback, and demonstrating teaching strategies, among other activities. It stands to reason that these kinds of activities would help teachers improve their practices, and thus help students make AYP. Yet, there is no evidence in our analysis that reading or other coaches (with one exception discussed below), are effective in promoting AYP related outcome measures. So, one interpretation is that reading coaches in urban, suburban and rural schools have a

negative effect imply that the help of coaches is limited or of negative value.

Another interpretation is that this negative effect is an oversimplification. Instead, this finding may indicate that schools that are struggling with AYP and related outcome measures were more likely to staff reading coaches to help improve AYP related outcome measures. Therefore, the analysis is an artifact of schools that are in danger of not making AYP measurement, are the very schools that employ more reading coaches. That is, the number of reading coaches in these schools skews the results toward indicating that reading coaches have a negative effect on making AYP measures.

At the very least, schools with reading coaches were still struggling to make their way to positive AYP outcome measures. Yet, other research studies have found significantly higher reading scores because of coaches (Marsh et al., 2010; Lockwood et al., 2010; the three Elementary School Journal studies (look up); Shidler, 2010; and Matsumuar, et al., 2013). This indicates that the relationship between teachers on the one hand, and coaches and specialists on the other hand, is one that should continue to be examined.

A different finding is that science coaches in suburban schools measured significantly positive in keeping schools off the mandatory improvement list. Science coaches may be exactly what suburban schools, perhaps more than other coaches.

Also of note is that "specialists" seem to have no effect, which may indicate that most of the help schools are receiving since the signing of the No Child Left Behind Act of 2001 is now in coaching (Koepke, 2017).

Organizing time - The length of a school day showed the most consistent pattern of having a positive effect when organizing time with three out of six cells: two in urban schools, one in rural schools, and none in suburban schools. Lengthening the school day may provide more learning opportunities for urban and rural students, while suburban students may already have ample opportunities for learning without lengthening the school day. Though lengthening the school day has not shown strong effects in other studies, our nationally representative, divided by location, does show a benefit to urban and rural schools.

An earlier start time shows a positive effect for urban schools make AYP measurements, and negative for helping suburban schools to do so. This is a very interesting finding, because it was also what Keller and her associates (2014) found. The later start time appears in both of these studies to be helpful to suburban students, but not urban students, which seem counter intuitive, and not what Keller expected to find.

Finally, block scheduling showed a negative effect for urban schools making AYP. Earlier block scheduling studies show mixed and no effect on achievement tests and grades.

Grouping teachers and students - Among the three constructs of school organization, student grouping also showed some interesting outcomes.

Grades subdivided into small groups such as "houses" or "families," showed a negative effect for staying out of the mandatory improvement category for urban and suburban schools. This is somewhat surprising, on the one hand, because there is an intuitive appeal to the idea that as teachers get to know their student better, they can help them achieve more. On the other hand, small schools are, in effect, schools subdivided into teams. In a 2009 memo, a spokesperson for the Gates Foundation wrote that, "Research shows that there is only half as much variation in student achievement between schools as there is among classrooms in the same school." (as cited in Strauss, 2014), and the Foundation decided to end pursuing its small school initiative, and focus on teaching and charter schools instead (Strauss, 2014).

Multi-age grouping shows a negative effect for suburban and rural schools staying out of the mandatory improvement category. This is a new finding as the multi-age grouping

literature has not shown much of an effect either way heretofore.

Finally, looping shows a positive effect for rural schools to stay off a mandatory improvement list for rural schools. This may be because rural schools utilize grouping more than urban or suburban school due to small class sizes. A few other studies found positive effects for looping (Franz et. al., 2010; Hill & Jones, 2018).

About the Authors

Sue Poppink is an associate professor at Western Michigan University with research interests in K-12 administration and policy. In addition, she has research interests in teaching, mentoring and identity development of practitioner-scholar doctoral students. Email: sue.poppink@wmich.edu

Xin Ma is a professor at the University of Kentucky with research interests in advanced statistical (quantitative) methods, advanced data analysis of large-scale (state, national, and international) surveys, psychology of mathematics education, program evaluation and policy analysis, and organizational (school) effectiveness and improvement. Email: xin.ma@uky.edu

Jianping Shen is the John E. Sandberg Professor of Education at Western Michigan University with research interests in leadership theory, data-informed decision-making, teacher retention and attrition, alternative certification, systemic change, and others. He uses both quantitative and qualitative methodologies. Email: jianping.shen@wmich.edu

References

Adelman, N. E., Haslam, M. B., & Pringle, B. A. (1996). *Studies of education reform: The uses of time for teaching and learning*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. Retrieved from ERIC (ED397562)

Bean, R. M., Draper, J. A., Hall, V., Vandermolen, J., & Zigmond, N. (2010). Coaches and coaching in Reading First schools: A reality check. *Elementary School Journal*, 111(1), 87–114. doi.org/10.1086/653471

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal*, 111(1), 7–34. doi.org/10.1086/653468

Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 111(3),* 430-454. doi.org/10.1086/657654.

Canady, R. L., & Rettig, M. D. (1996). Block scheduling: What is it? Why do it? How do we harness its potential to improve teaching and learning? In R. L. Canady & M. D. Rettig (Eds.), *Teaching on the block* (pp. 1-27). Larchmont, NY: Eye on Education.

Carskadon, M. A., Wolfson, A. R., Tzischinsky, O., & Acebo, C. (1995). Early school schedules modify adolescent sleepiness. *Sleep Research, 24*(92).

Cascio, E. U., & Schanzenbach, D. W. (2016). The impacts of expanding access to high-quality preschool education. Brookings Paper on Economic Activity, the Brookings Institution. 47(2). 127-192. doi: 10.3386/w19735

Cooper, H., Nye, B., Charlton, K., Lindsey, J. J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227-268. doi: 10.3102/00346543066003227

Cooper, H., Valentine, J. C., Charlton, K., & Melson, A. (2003). The effects of modified school calendars on student achievement and on school and community attitude. *Review of Educational Research, 73*(1), 1-52. doi: 10.3102/00346543073001001

Farbman, D., Christie, K., Davis, J., Griffith, M., & Zinth, J. D. (2011). *Learning time in America: Trends to reform the American school calendar*. Denver, CO: Education Commission of the States, and Boston, MA: National Center on Time & Learning.

Folsom, J. S., Osborne-Lamkin, L., Cooley, S., & Smith, K. (2017). *Implementing the extended school day policy in Florida's 300 lowest performing elementary schools* (REL 2017-253). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational laboratory Southeast.

Folsom, J. S., Petscher, Y., Osborne-Lampkin, L., Cooley, S., Herrera, S., Partridge, M., & Smith, K. (2016). *School reading performance and the extended school day policy in Florida* (REL 2016–141). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Franz, D. P., Thompson, N. L., Fuller, B., Hare, R. D., Miller, N. C., & Walker, J. (2010). Evaluating mathematics achievement of middle school students in a looping environment. *School Science and Mathematics Association, Inc., 110*(6), 298-308. doi:10.1111/j.1949-8594.2010.00038.x

Fredrickson, J., Wahlstrom, K., & Wrobel, G. (1999). *The adolescent crisis: Can later school start times help?* Paper presented at the American Educational Research Association, Montreal, Canada.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. Washington, DC: U.S. Department of Education.

Goodlad, J. I., & Anderson, R. H. (1959). *The nongraded elementary school*. New York: Harcourt, Brace & World.

Gromada, A., & Shewbridge C. (2016). *Student learning time: A literature review* (OECD Education Working Papers, no. 127). Paris, France: OECD Publishing. doi.org/10.1787/5jm409kqqkjh-en

Graves, J. (2011). Effects of year-round schooling on disadvantaged students and the distribution of standardized test performance. *Economics of Education Review, 30.* 1281-1305. doi.org/10.1016/j.econedurev.2011.04.003

Graves, J., McMullen, S., & Rouse, K. (2013). Multi-track year-round schooling as cost saving reform: Not just a matter of time. *Education Finance and Policy, 8(3),* 300-315. doi: 10.1162/EDFP_a_00097

Gutierrez, R., & Slavin, R. (1992). Achievement effects of the nongraded elementary school: A best evidence synthesis. *Review of Educational Research, 62*(4), 333-376. doi: 10.3102/00346543062004333

Hill, A. J., & Jones, D. B. (2018). A teacher who knows me: The academic benefits of repeat student-teacher matches. *Economics of Education Review, 64*, 1-12. doi.org/10.1016/j.econedurev.2018.03.004

Kallery & Loupidou. (2016). Learning science in small multi-age groups: The role of age composition. *International Journal of Science Education. 38(9),* doi.org/10.1080/09500693.2016.1201871

Keller, P. S., Smith, O. A., Gilbert, L. R., Bi, S., Haak, E. A. & Buckhalt, J. A. (2014). Earlier schools start times as a risk factor for poor school performance: An examination of public elementary schools in the Commonwealth of Kentucky. *American Psychological Association*, 107, doi.org/10.1037/a0037195

Koepke, J. (2017). Literature review: Literacy coaching. MidSouth Literacy Journal, 2(2), (24-30).

Lockwood, J. R., McCombs, J. S., & Marsh, J. (2010). Linking reading coaches and student achievement: Evidence from Florida middle schools. *Educational Evaluation and Policy Analysis, 32*(3), 372-388. doi: 10.3102/0162373710373388

Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision-making: Policy implementation and effects in Florida middle schools. *Educational Policy, 24*(6), 872-907. doi: 10.1177/089504809341467

Matsumura, L. C., Garnier, H. E., Correnti, R., Junker, B., & Bickel, D. D. (2010). Investigating the effectiveness of a comprehensive literacy coaching program in schools with high teacher mobility. *Elementary School Journal*, 111(1), 35–62. doi.org/10.1086/653469

Matsumura, L. C., Garnier, H.E., & Spybrook, J. (2013). Literacy coaching to improve student reading achievement: A multi-level mediation model. *Learning and Instruction, 25,* 35-48. doi.org/10.1016/j.learninstruc.2012.11.001

National Association for Year-Round Education. (2000). *Twenty-seventh reference directory of year-round education programs for the 2000-2001 School Year*. San Diego, CA: Author.

Patall, E. A., Cooper, H., & Allen, A. B. (2010). Extending the school day or school year: A systematic review of research (1985 - 2009). *Review of Educational Research, 80*(3), 401-436. doi: 10.3102/0034654310377086

Quatroche, D. J., Bean, R. M., & Hamilton, R. L. (2001). The role of the reading specialist: A review of the research. *The Reading Teacher, 55*(3), 282-294.

Ross, J. A. (1992). Teacher efficacy and the effects of coaching on student achievement. *Canadian Journal of Education, 17*(1), 51-65. doi: 10.2307/1495395

Sailors, M., & Shanklin, N. L. (2010). Introduction: Growing evidence to support coaching in literary and mathematics. *The Elementary School Journal, 111*(1), 1-6. doi: 10.1086/653467

Shidler, L. (2009). The impact of time spent coaching for teacher efficacy on student achievement. *Early Childhood Education Journal, 36*, 453-460. doi: 10.1007/s10643-008-0298-4

Stengel, R. (2009, April 15). Arne Duncan: The apostle of reform. *TIME, Retrieved from http://www.time.com*.

Strauss, V. (2014, June 9). How much Bill Gates's disappointing small-school effort really cost. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2014/06/09/how-much-bill-gatess-disappointing-small-schools-effort-really-cost/?utm_term=.fbf905bb64ed

Texas Education Agency. (1999). *Block scheduling in Texas public high schools*. Austin, TX: Author.

Thum, Y. M., & Hauser, C. H. (2015). *NWEA 2015 MAP Norms for Student and School Achievement Status and Growth.* NWEA Research Report. Portland, OR: NWEA

Tyack, D. B. (1974). *The one best system: A history of American urban education.* Cambridge, MA: Harvard University Press.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research, 65*(4), 319-381. doi: 10.3102/00346543065004319

Wahlstrom, K. (1999). *Adolescent sleep needs and school starting times.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Wahlstrom, K. (2002). Changing times: Findings from the first longitudinal study of later high school start times. *National Association of Secondary School Principals (NASSP) Bulletin, 86*(633), 3-21. doi: 10.1177/019263650208663302

Wahlstrom, K., & Freeman, C. M. (1997). *School start time study: Preliminary report of findings*. Minneapolis, MN: University of Minnesota: Center for Applied Research and Educational Improvement.

Wheeler, P. (1987). The relationship between grade six test scores and the length of the school day. *Educational Research Quarterly, 11*(3), 10–17.

Woods, J. R. (2015, August). Instructional time trends. *Education Trends.* Denver, CO: Education Commission of the States. Retrieved from: https://www.ecs.org/

Zepeda, S. J., & Mayers, R. S. (2006). An analysis of research on block scheduling. *Review of Educational Research, 76*(1), 137-170. doi: 10.3102/00346543076001137

Appendix A

Descriptions of Outcome, Control, and Predictor Variables

Outcome variables
Schools make AYP is a dummy variable based on the following item: Adequate Yearly Progress (AYP) is your state's measure of yearly progress toward achieving state academic standards. At the end of the LAST school year (2006-2007), did this school make Adequate Yearly Progress (AYP)? (Yes, No)
Schools are unidentified for improvement is a dummy variable based on the following item: A school is identified for improvement if it does not make Adequate Yearly Progress (AYP) for two consecutive years or more in the same content area. At the end of the LAST school year (2006-2007), was this school identified for improvement due to Adequate Yearly Progress (AYP) requirements? (Yes, No)
Control variables
Proportion of students eligible for free or reduced-price lunch is a continue variable in the form of percent (for each school).
Proportion of minority students is a continue variable in the form of percent (for each school).
School type is three dummy variables [middle school, high school, and combine (middle and high) school] with elementary school as the reference.
School location is used for stratification of schools into urban, suburban, and rural schools.
Predictor variables
Length of a school day is a continue variable based on the following item: How long is the school day for students in this school?
Time a school day begins is a continue variable based on the following item: At what time do most of the students in this school begin the school day?
School days per year is a continue variable based on the following item: How many days are in the school year for students in this school?
Traditional grades or academic discipline-based departments, g*rades subdivided into small groups*, *student groups with the same teacher two or more years*, *multi-age grouping*, and *block scheduling* are dummy variables based on the following item: THIS school year (2007-08), does this school use the following methods to organize most classes or most students? Traditional grades or academic discipline-based departments? Grades subdivided into small groups such as "houses" or "families?" Student groups that remain two or more years with the same teacher? Multi-age grouping? Block scheduling? (Yes, No)

A year-round calendar to distribute school days across 12 months is a dummy variable based on the following item:

Does this school use a year-round calendar to distribute school days across 12 months? (Yes, No)

Reading specialist, *math specialist*, and *science specialist* are dummy variables based on the following item:

Do any of the teachers or staff have the following specialist assignments in this school? (A specialist works with students.) Reading specialist? Math specialist? Science specialist? (Yes, No)

Reading coach, *math coach*, and *science coach* are dummy variables based on the following item:

Do any of the teachers or staff have the following coaching assignments in this school? (A coach works with teachers. Coaching includes observing lessons, providing feedback, and demonstrating teaching strategies.) Reading coach? Math coach? Science coach? (Yes, No)

Appendix B

Descriptive Statistics of Outcome, Control, and Predictor Variables by School Location

	Urban schools		Suburban schools		Rural schools	
	Mean	SD	Mean	SD	Mean	SD
Outcome variables						
Schools make AYP (coded as 1)	.58	.49	.71	.45	.76	.43
Schools are unidentified for improvement (coded as 1)	.74	.44	.87	.34	.88	.32
Control variables						
School size	7.49	6.48	7.50	5.85	4.19	3.26
Percent of students eligible for free or reduced-price lunch	5.06	3.08	3.47	2.68	4.67	2.63
Percent of minority students	5.91	3.26	3.31	3.12	2.71	3.13
Middle school (versus elementary school)	.14	.35	.16	.36	.11	.31
High school (versus elementary school)	.32	.47	.37	.48	.34	.47
Combined (middle and high) school (versus elementary school)	.08	.28	.08	.27	.21	.41
Predictor variables						
Length of a school day (in hours)	6.70	.66	6.70	.54	6.84	.55
Time a school day begins (in measurement unit of 10 minutes)	10.12	3.08	9.74	2.95	9.83	2.05
School days per year (in measurement unit of 10 days)	17.98	.59	17.89	.47	17.74	.64
Traditional grades or academic discipline-based departments	.93	.25	.94	.24	.95	.21
Grades subdivided into small groups	.23	.42	.18	.38	.11	.31
Student groups with the same teacher two or more years	.17	.37	.13	.34	.14	.35
Multi-age grouping	.21	.41	.20	.40	.24	.43

Block scheduling	.36	.48	.33	.47	.28	.45
A year-round calendar to distribute school days across 12 months	.17	.38	.14	.35	.12	.32
Reading specialist	.55	.50	.53	.50	.42	.49
Math specialist	.24	.43	.17	.38	.15	.36
Science specialist	.12	.32	.07	.25	.06	.24
Reading coach	.39	.49	.28	.45	.22	.41
Math coach	.25	.43	.14	.35	.10	.30
Science coach	.09	.29	.06	.25	.04	.19

Note. SD = standard deviation. Sample size for urban schools is 1,670. Sample size for suburban schools is 3,220. Sample size for rural schools is 2,570. In SASS, school size is a collapsed index of total enrollment in school. To measure time a school day begins, we count the time from 6 o'clock in the morning.