

*American Educational Research Journal*  
Month XXXX, Vol. XX, No. X, pp. 1–32  
DOI: 10.3102/0002831209350494  
© 2009 AERA. <http://aerj.aera.net>

# Evaluating the Relationship Between Student Attendance and Achievement in Urban Elementary and Middle Schools: An Instrumental Variables Approach

Michael A. Gottfried  
*University of Pennsylvania*

*Researchers, policymakers, practitioners, and parents have assumed a positive relationship between school attendance and academic success. And yet, among the vast body of empirical research examining how input factors relate to academic outcomes, few investigations have honed in on the precision of the relationship between individual attendance and student achievement. The purpose of this article is to provide insight into this relationship. Specifically, this study has evaluated the hypothesis that the number of days a student was present in school positively affected learning outcomes. To assess this, a unique empirical approach was taken in order to evaluate a comprehensive dataset of elementary and middle school students in the Philadelphia School District. Employing a fixed effects framework and instrumental variables strategy, this study provides evidence from a quasi-experimental design geared at estimating the causal impact of attendance on multiple measures of achievement, including GPA and standardized reading and math test performance. The results consistently indicate positive and statistically significant relationships between student attendance and academic achievement for both elementary and middle school students.*

**KEYWORDS:** achievement, attendance, urban education

Though the empirical evaluation of the relationship between individual attendance and student-level achievement has received little attention among education researchers (Corville-Smith, 1995; Epstein & Sheldon,

---

MICHAEL A. GOTTFRIED is completing his PhD in the Applied Economics Group in The Wharton School of Business, University of Pennsylvania, 1400 Steinberg Hall-Dietrich Hall, 3620 Locust Walk, Philadelphia, PA 19104; e-mail: [mgottfr2@wharton.upenn.edu](mailto:mgottfr2@wharton.upenn.edu). He is also an Institute of Education Sciences Predoctoral Fellow. By June 2010, he will have joined RAND Corporation, Santa Monica, California, as an associate policy researcher. His research focuses on urban education issues pertaining to student outcomes, peer effects, school governance, and district reform.

2002), attendance is nonetheless credited as being an important component of school success. Students with better attendance records are cited as having stronger test performance (Balfanz & Byrnes, 2006; Lamdin, 1996; Nichols, 2003). Other research has supported the notion that student attendance records can serve as direct signals of school quality (Coutts, 1998). Several studies have deemed attendance as important enough to be evaluated as an academic outcome (King, 2000; Lehr, Sinclair, & Christenson, 2004; Phillips, 1997; Sheldon, 2007), thereby suggesting that increased attendance is a direct indicator, rather than determinant, of school success.

Most research, however, has focused on negative issues pertaining to decreased attendance. Lower attendance rates have been cited as detrimental to learning and academic achievement, and an increase in absences in elementary and middle school can be predictive of higher risk factors in both concurrent and future years of education (Dreyfoos, 1990; Finn, 1993; Lehr et al., 2004; Stouthamer-Loeber & Loeber, 1988). Academically, students who do not attend school as frequently receive fewer hours of classroom instruction and consequently perform more poorly on exams in that same year (Chen & Stevenson, 1995; Connell, Spencer, & Aber, 1994; Finn, 1993; Nichols, 2003). Having consistently low levels of attendance in early grade levels is also correlated with higher future academic risks, including nonpromotion (Neild & Balfanz, 2006) and dropping out (Rumberger, 1995; Rumberger & Thomas, 2000).

There are also sociological and economic concerns associated with students having low attendance rates. Sociologically, decreased attendance is related to increased alienation from classmates, teachers, and schools (Ekstrom, Goertz, Pollack, & Rock, 1986; Finn, 1989; Johnson, 2005; Newmann, 1981). Missing school is also correlated with current and future risky behaviors, such as tobacco, alcohol, and drug use (Halfors et al., 2002; Wang, Blomberg, & Li, 2005). Economically, students who do not attend school as frequently (and thus have a higher correlated risk for nonpromotion and dropping out as mentioned above) tend to face greater future financial hardships, such as unemployment (Alexander, Entwisle, & Horsey, 1997; Broadhurst, Patron, & May-Chahal, 2005; Kane, 2006).

Research has suggested that the academic, sociological, and economic issues related to both increased and decreased attendance are heightened for youth in urban school systems, such as the Philadelphia School District (Balfanz & Legters, 2004; Fine, 1994; Orfield & Kornhaber, 2001). For instance, increased attendance in math classes has been attributed with reducing the severity of the math achievement gap for urban students (Balfanz & Byrnes, 2006). Thus, the importance of attending school in early years appears to be crucial for urban youth, because it is particularly these minority and high-poverty students who fall behind in math achievement beginning as early as fourth grade (Balfanz & Byrnes, 2006). If the premise is true—that increasing school attendance influences the acquisition of

### *Evaluating the Attendance-Achievement Relationship*

foundational math skills in elementary grades—then identifying early strong relationships between attendance and achievement can be indicative of reduced future academic risk among young urban students.

On the other hand, decreased attendance is correlated with exacerbated academic issues for urban, minority youth, especially when compared to their non-urban, non-minority counterparts (Orfield, Losen, Wald, & Swanson, 2004; Swanson, 2004). Further, declining levels of attendance are related to deteriorating academic outcomes for these urban students as they progress into later grades of schooling (Fine, 1991; Wasley, 2002). For instance, Easton and Englehard (1982) found that within an urban school district, student absences were negatively correlated with reading achievement, and this relationship became even stronger as students entered grades 7 and 8. They attributed this heightened negative relationship to the fact that family home environments became less important in academic development compared to school settings for urban middle school students.

By using quasi-experimental methods to explore the precision of the relationship by which student attendance is related to both grade point average (GPA) and standardized testing outcomes, this article develops a deeper empirical understanding of the extent to which increased attendance is related to multiple measures of school success. Doing so is particularly useful for a population of urban minority youth, who are especially vulnerable to issues surrounding increased or decreased attendance. Therefore, if the analysis to follow can suggest causal effects of attendance on academic performance, then the results would yield further insight into the early educational experiences of urban youth, before they enter into high school where the probability of school failure or behavior problems becomes intensified (Alexander et al., 1997; Barrington & Hendricks, 1989; Lehr et al., 2004).

### **Empirical Background**

Of the empirical studies that have directly focused on the relationship between attendance and achievement, most have been conducted at the university level (Devadoss & Foltz, 1996; Durden & Ellis, 1995; Kirby & McElroy, 2003; Marburger, 2001). As an example, Levine (1992) found that in psychology college courses, weaker class attendance policies led to negative correlations between absences and class averages. In a similar study on the role of undergraduate attendance on achievement, Romer (1993) found that the grades of regular attendees in economics courses were a full letter higher than those grades of sporadic attendees. On the basis of this finding, Romer suggested that policies aimed at increasing attendance in college, particularly those making attendance mandatory, might be considered.

In the scant quantitative literature that does exist surrounding K–12 educational issues of attendance and achievement, the results are mixed. Caldas

(1993) found that attendance was positively and significantly related to student performance in Louisiana's public elementary and secondary schools. His unit of observation was at the school level, and attendance was quantified as the percent of students per school who were present in a given day. As a result, the conclusions from the study are grounded in the relationship between the average rate of school attendance and average student performance. Thus, findings from this article cannot explain the relationship between individual attendance and student achievement.

Lamdin (1996) also relied on aggregate data to show that student attendance had a positive and significant relationship with academic performance. As with the Caldas (1993) study, the results derived from this article can only be interpreted in terms of average school performance. In fact, the author asserted that conclusions regarding individual-level achievement outcomes cannot be drawn from this study. Borland and Howsen (1998) refuted Lamdin's (1996) findings, pointing to the absence of a measure of ability within the analysis. The authors deemed this to be problematic because unmeasured ability could be correlated with both attendance and achievement and hence could have confounded Lamdin's (1996) results. The authors did find that once ability was incorporated into the model, attendance became a non-significant predictor of student achievement. Their refutation was based on aggregated variables as well.

More recently, Roby (2004) concluded that based on the analysis of educational outcomes in Ohio, there was a statistically significant relationship between attendance and achievement in 4th, 6th, 9th, and 12th grades. Sheldon (2007) also used Ohio testing data and evaluated the results of attendance interventions in 2001. His analyses indicated that schools with high rates of daily attendance were positively correlated with student performance on reading and math achievement tests. Like previous research, these two studies have employed measures of attendance only at an aggregated level rather than for individual students.

This article will address several gaps in the empirical literature. First, those studies that have attempted to hone in on the relationship between student attendance and achievement in K-12 education have been conducted only at an aggregated level of analysis. Although these articles have provided insight into how attendance is related to achievement in a student's educational experience, aggregate data have less variability than do underlying individual-level data. As a result, it is not possible with aggregate data to offer empirical findings on the relationship between student-level attendance and student-level achievement. More importantly, relying on aggregate measures of attendance to make claims about individual student behavior is subject to ecological fallacy, in which inferences of students would be based on school-wide metrics. Thus, this article extends upon those previous empirical studies, which employed only aggregate data, with new analyses at the student level. By using a large-scale, longitudinal

### *Evaluating the Attendance-Achievement Relationship*

database of individual- and multi-level observations for elementary and middle students in the Philadelphia School District from academic years 1994/1995 through 2000/2001, this article will provide insight into how individual attendance relates to student performance at a more detailed level of analysis.

Second, as Borland and Howsen (1998) have suggested, there is need for quantifying a measure of ability when conducting analyses that model student achievement as the dependent variable. Because ability may be positively correlated with attendance, there are confounding effects between both attendance and achievement that may bias the estimates on the number of days a student is present at school. By using a longitudinal dataset that has multiple observations of students over time, this study has minimized this issue by implementing a value-added model of student achievement. As further explained in the Method section below, the feature of including a lagged achievement score at the individual level means that, under the assumptions of the model, it is no longer necessary to incorporate additional measures of ability or a full historical panel of information on any particular student. Thus, in this article, student ability no longer confounds the relationship between attendance and achievement.

Finally, there is an additional issue regarding confounding variables that the literature has not entirely addressed. Even after controlling for student ability either through a contemporaneous specification or value-added model, it is still possible that there are unobserved factors affecting measures of both student attendance and academic outcomes. For instance, it might be hypothesized that unobserved student motivation or family environments can simultaneously influence both attendance and achievement. As a result, the coefficients from ordinary least squares regressions would remain biased. This article attempts to rectify this confounding issue by implementing an instrumental variables strategy. That is, if it were possible to find a measure that embodies an exogenous source of variation affecting only student attendance but not achievement, then this quasi-experimental approach would be appropriate (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007).

This article claims to have found such a variable, hence allowing for the employment of the instrumental variables approach. The instrument utilized in this study is the geographical distance that a student lives from school. This instrument has appeal because attendance tends to decline as a student's mileage from school increases, as this article will demonstrate. Previous studies on this relationship between geographic distance and school attendance confirm this relationship both at the K-12 level (Jensen & Nielsen, 1997; Schlossberg, Greene, Phillips, Johnson, & Parker, 2006; Schultz, 2004) and for college students (Frenette, 2004). On the other hand, it is not evident that the distance a student lives from school has a direct relationship to GPA or standardized achievement. Moreover, a student's distance from

school is related only to his or her attendance and not the attendance of teachers. As such, this study has eliminated an often confounding issue, in which a factor that affects student attendance is also influencing teacher attendance (or vice-versa), thereby biasing estimates. Rather, the relationship between distance from school and measures of academic achievement is only mediated through the student. With this exogenous source of variation, the methodology presented in this article allows for the estimation of a causal relationship between attendance and achievement.

## Data

The analysis of attendance and student achievement in this study utilizes a comprehensive dataset of student and neighborhood observations in the Philadelphia School District. This study implements data from all elementary and middle schools in the Philadelphia School District between the academic years 1994/1995 through 2000/2001. Over this time period, this amounts to 223 comprehensive, neighborhood elementary and middle schools, with approximately  $N = 86,000$  students in kindergarten through grade 8. In total, there are  $N = 332,000$  student-year observations.<sup>1</sup>

Student data were obtained from the School District of Philadelphia via the District's Office of Student Records and District's Personnel Office. Neighborhood data were obtained from the 2000 Census flat files at the census block level. Neighborhood data relating to age, sex, households, families, and housing units were merged from the Census Summary File 1; additional social, economic, and housing measures were merged from Summary File 3. The data sample in Summary File 3 includes one in six households that received the long-form Census survey, whereas Summary File 1 measures are based on the full universe of responding households.

Every student in the database can be tracked by teacher, classroom,<sup>2</sup> grade, and year throughout his or her tenure in the Philadelphia School District. Students cannot be tracked if they leave the school system: No information is retained for those students who exit the Philadelphia School District for other districts, private, or parochial schools.<sup>3</sup> However, because students keep their unique identification number in the District's record system, if students should return into the Philadelphia School District, they can be matched back to their original record. Because of this intricate tracking mechanism of incoming, outgoing, and returning students, the sample includes the entire population of students per grade in the Philadelphia public school system over the period from years 1994/1995 to 2000/2001.

Table 1 presents descriptive statistics for those students in the dataset. The table provides information on the overall sample as well as elementary and middle school subsamples. There are more observations for elementary school students than for middle school students. This results from the construction of the database and specifically how the data on elementary and

*Evaluating the Attendance-Achievement Relationship*

**Table 1**  
**Descriptive Statistics for All Students, Academic Years**  
**1994/1995 Through 2000/2001**

	Full Sample		Elementary		Middle	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>N</i>	332,924		293,449		39,475	
Academic outcome measure						
GPA	2.44	1.10	2.44	1.10	2.44	1.08
Attendance measures						
Total days present	165.26	17.20	165.64	16.98	162.40	18.51
Distance from school (in miles)	1.65	2.08	1.66	2.10	1.54	1.93
Student characteristics (%)						
Male	49.26	49.99	49.44	50.00	47.90	49.96
White	17.96	38.38	17.77	38.23	19.34	39.50
Black	67.51	46.83	67.98	46.66	64.01	48.00
Latino	10.73	30.95	10.66	30.86	11.26	31.61
Asian	3.60	18.64	3.39	18.11	5.16	22.13
Other	0.20	4.47	0.20	4.44	0.22	4.66
Special education	0.58	7.57	0.50	7.02	1.19	10.82
Free lunch recipient	61.39	48.69	61.73	48.60	58.88	49.21
English language learner	4.38	20.46	4.21	20.08	5.58	22.96
Behavior problem	10.41	30.54	10.75	30.97	8.73	28.23
Student's neighborhood block characteristics						
Percent of block, White	30.56	32.95	30.15	32.93	33.64	32.94
Percent of block, below poverty	14.16	8.71	14.22	8.72	13.68	8.65
Household vacancy rate for block	12.75	9.32	12.81	9.32	12.28	9.31
Log of average block income	10.16	0.45	10.15	0.45	10.19	0.46

middle school students were collected. The student data were organized into five cohorts of students, observed between 1994/1995 through 2000/2001. Cohorts 1, 2, and 3 were in kindergarten, first, and second grades, respectively, in 1994. Cohort 4 was in kindergarten in 1995 and cohort 5 in 1996. As such, by the academic year 2000, only three cohorts will have reached middle school. Thus, there is not an equal number of elementary and middle school observations within and between cohorts.

The primary outcome variable used in this study was GPA. Although GPA is a measure of academic performance and has important implications, it can be subject to differing interpretations. For instance, it is possible that GPA reflects attendance as well as academic performance (Marburger, 2006). Further, teachers might decrease expectations for low socioeconomic status (SES) students and therefore adjust their grading schemes accordingly,

often resulting in lower grading outcomes (Van Matre, Valentine, & Cooper, 2000). As such, GPA may not serve as an absolute measure of achievement, but it nonetheless provides an indication of academic attainment, and the fact that the grades are retained on a student's record certainly bears weight on future decisions made by both students and school administrators. Thus, higher GPAs certainly indicate in some fashion better school success and predict stronger future educational outcomes, such as lower rates of dropping out and higher college enrollment (Anaya, 1999; Kuncel, Crede, & Thomas, 2005; Rumberger & Palardy, 2005; Willingham & Breland, 1982).<sup>4</sup> For comparability, a second analysis was conducted on third and fourth grade students for whom the Stanford Achievement Test Ninth Edition (SAT9) reading and math scores were available. Generalizing the analysis beyond GPA, the use of standardized test outcomes allows for the interpretation of the prediction of attendance based on both teacher-assigned and standardized measures of academic performance.

Table 1 shows that GPA is fairly normally distributed with a mean hovering around 2.4 for all groups (with A being a 4 and F being a 0). Additionally, because of the longitudinal nature of this dataset, using GPA as a dependent variable allows for a year-to-year evaluation of student performance and thus provides for a gauge of schooling attainment recorded at regularly timed intervals.<sup>5</sup>

There are additional student-level covariates implemented in this study. Attendance is measured as the total days a student is present in a given school year. To derive the distance in miles that a student lives from school, which will be used in conjunction with attendance in the instrumental variables section of this article, information was collected on a student's home address, including street number and name and zip code. The merging of neighborhood data with a student's record was achieved by a geo-coding process in which student addresses were matched to their exact longitude and latitude coordinates. Using geo-coding has allowed for the determination of the exact geographic distance that students lived from the schools they attended. Approximately, 99% of the students in the Philadelphia School District lived within 10 miles of the school in which they were enrolled. Therefore, 10 miles was the cut-off distance for the analysis, as any observation with a higher mileage to school was determined to be a statistical outlier. Furthermore, the geo-coding process also enabled the assignment of each student to a census block group. Doing so provides information for every student on the percent of their census blocks that were White, the percent that were at or below poverty, vacancy rates, as well as the average income of each block. Table 1 demonstrates that these measures are relatively similar for both elementary and middle school samples.

For every student in each academic year, the dataset contains demographic information concerning personal characteristics, such as gender

and race. In Table 1, there are higher percentages of White and Asian students in middle schools compared to elementary schools, whereas the percentage of Black students is slightly lower for the middle school sample. Student demographic variables also include the following: an indicator for special education status; English language learner status; recipient of free lunch;<sup>6</sup> and whether or not the student has a behavior problem, determined by his or her behavior grade from the previous year's record. There tends to be more students identified as having special education status in middle school, potentially because these needs might not be immediately recognized in primary years of schooling.

In both elementary and middle schools in Philadelphia, almost 60% of the students would be classified as low SES, as determined by receiving free lunch. In the absence of family information, free lunch status and neighborhood information often serve in empirical models as proxies for family background (e.g., Hanushek, Kain, Markman, & Rivkin, 2003), as they are based on direct observation of family and neighborhood characteristics (e.g., household and census block incomes).

## Method

### A Baseline Model

In the baseline empirical model of educational outcomes, student achievement can be described using a linear relationship with a particular academic measure as the dependent variable and a vector of independent variables. In the main analyses of this study, student GPA is the dependent variable and total days present, student demographics, and neighborhood characteristics are independent variables. This baseline model is displayed in Equation 1:

$$Y_{ijgkt} = \beta_0 + \beta_1 PR_{it} + \beta_2 S_{it} + \beta_3 N_{it} + \gamma_k + \omega_g + \nu_t + \varepsilon_{ijgkt} \quad (1)$$

where  $Y$  is the GPA for student  $i$  in classroom (or homeroom)  $j$  in grade  $g$  in school  $k$  in year  $t$ ;  $PR$  is student  $i$ 's attendance information in year  $t$ ;  $S$  is a vector of student-level characteristics in year  $t$ ; and  $N$  includes neighborhood characteristics for student  $i$  in year  $t$ .

The final four terms of Equation 1 delineate the fixed effects selected in this empirical specification as well as the error term. In particular,  $(\gamma_k)$  are school fixed effects,  $(\omega_g)$  are grade fixed effects,  $(\nu_t)$  are year fixed effects, and  $(\varepsilon_{ijgkt})$  is a random error capturing individual variations over time as well as a class-specific random component that is common to all members of the same classroom or homeroom. Empirically, the final component of Equation 1 is estimated with Huber/White/sandwich standard errors that are adjusted for classroom clustering.

School fixed effects ( $\gamma_k$ ) control for the influences of the school-level environment by capturing systematic differences across each unique institution. By, in essence, holding constant those time-invariant school-specific characteristics, such as curriculum, school neighborhood, leadership, organization, and hiring practices, the empirical model controls for the educational effectiveness at the school level, thereby allowing for the evaluation of within, rather than between, school variation. Analogously, ( $\omega_g$ ) and ( $v_t$ ) are control variables for academic grades and calendar years, respectively, hence enabling the analysis of the model to parse out the effects of schools from other common influences that occur across the population for a given cohort and in a given year.<sup>7</sup>

### Strengthening the Baseline Model

Estimates of  $\beta_1$ , the coefficient on total days present in a given school year, may be biased under ordinary least squares regressions, even with the fixed effects and error structure designated in Equation 1. The reason is that there are current and past unobservable factors that can influence both student attendance as well as GPA. For instance, family environments or student motivation might simultaneously be related to both days present and GPA. As a first attempt to remedy this problem, a value-added model strategy is employed to mitigate this bias, under the assumption that unobservable student-level influences of achievement are time-invariant. In a value-added model, one-year lagged GPA serves as a proxy for individual student fixed effects. In this case, if the attributes of a student's family environment are consistent over time, then this model would more accurately estimate the relationship between attendance and achievement than would the baseline model.

The motivation for this strategy lies behind the construction of the value-added specification, which begins with a historical model of academic achievement. Conceptually, a historical model assumes that current achievement is a function of all current and prior inputs to schooling. The linear historical model of achievement in Equation 1 would be expanded to incorporate all time periods  $t$  through  $(t - n)$ , where  $(t - n)$  is the period in which the student initially enters school. However, it is difficult to acquire all inputs to estimate an equation of this sort that spans a student's educational history. One solution to this problem is to take the difference of the historical model of Equation 1 with respect to year  $t$ , the current year of schooling, and the historical model with respect to year  $t - 1$ , the previous year of schooling. The result is known as the value-added specification, where all input requirements reduce to current inputs plus achievement from the  $t - 1$  period:

$$Y_{ijgkt} = \beta_0 + \beta_1 PR_{it} + \beta_2 S_{it} + \beta_3 N_{it} + \beta_4 Y_{ijgk(t-1)} + \gamma_k + \omega_g + v_t + \varepsilon_{ijgkt} \quad (2)$$

As noted by the subscripts, this value-added model strictly incorporates current achievement as the dependent variable and previous achievement

and contemporaneous covariates as input variables.<sup>8</sup> Importantly, the result of subtracting the historical model of Equation 1 with respect to  $t - 1$  from a model with respect to  $t$  is that it removes the empirical requirement of having to directly estimate historical measured and unmeasured influences on student achievement. Rather, because the value-added model retains a measure of lagged achievement, this covariate is assumed to capture the influences of historical inputs for student  $i$ , thereby leaving only current measures to be estimated. Thus, biases that were created by omitted historical variables only bias the estimated coefficient of lagged achievement (Hanushek et al., 2003; Zimmer & Toma, 2000). Since this lagged variable accounts for historical information about the student, this measure can in essence serve as a proxy for individual fixed effects.<sup>9</sup>

Even with the use of a proxy for individual fixed effects (or true individual fixed effects as conducted in alternative specifications of this model) as well as the school, year, and grade fixed effects, these measures are nonetheless constructed under the assumption that unobserved variables are time-invariant. However, there can also be unobservable factors that are time-variant, and the use of implementing fixed effects would not necessarily remedy this problem (Miller, Murnane, & Willett, 2008). As an example, the estimation of the baseline or value-added models does not allow for reduction in biases resulting from concurrent unobservable influences in year  $t$ , such as this year's unique effort level or motivation. Thus, despite the use of a lag and other fixed effects, the relationship between attendance and achievement may still reflect the impact of omitted factors on regressors. It is necessary to turn to estimation strategies that are more immune to this time-variant omitted variable bias. In this study, in conjunction with baseline and value-added models, an instrumental variables strategy is employed as a more refined attempt of removing the bias on  $\beta_1$ .

Implementing the instrumental variables strategy requires a two stage least squares format, in which there is a unique regression equation for each stage. Rather than immediately evaluating the relationship between days present and GPA, this approach begins with a first stage of analysis. As previously mentioned, the reason that this first stage is necessary is because unobserved, time-variant influences may affect both independent and dependent variables. If so, the results would yield biased estimates. However, in the first stage,  $PR_{it}$  becomes the outcome variable. The independent variables include all covariates that will be used in the second stage as well as an instrument. The instrument, or exogenous independent variable, is distance from school, and it must not be directly correlated with GPA, except through its relationship with student attendance. As described in Table 2, the correlation between distance and GPA is  $-0.05$  and between distance and lagged GPA is  $-0.03$ . On the other hand, the correlation between distance and attendance is  $-0.59$  for the full sample. The first stage equation looks as follows:

Table 2  
Correlation Coefficients Relating to Mileage Distance From School

	Full Sample	Elementary	Middle
Days present	-0.587***	-0.615***	-0.433***
GPA	-0.049***	-0.049***	-0.044***
GPA, lagged	-0.040***	-0.040***	-0.030***
Male	-0.002	-0.001***	-0.010
White	-0.093***	-0.092***	-0.088***
Black	0.130***	0.124***	0.162***
Hispanic	-0.055***	-0.053***	-0.073***
Asian	-0.045***	-0.038***	-0.083***
Other	0.005	0.005**	0.008
Special education	-0.005**	-0.006**	0.004
Free lunch recipient	0.048***	0.049***	0.028***
English language learner	-0.040***	-0.037***	-0.062***
Behavior problem	0.032***	0.032***	0.023***
Census block: percent White	0.021***	0.040***	-0.129***
Census block: percent poverty	-0.041***	-0.052***	0.048***
Census block: vacancy rate	-0.024***	-0.035***	0.059***
Census block: log(mean income)	0.052***	0.066***	-0.047***

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

$$PR_{it} = \eta_0 + \eta_1 D_{it} + \eta_2 X + \gamma_k + \omega_g + v_t + \varepsilon_{ijgkt} \quad (3)$$

where  $D_{it}$  is the instrumental variable (distance),  $X$  is a vector of all covariates from Equation 1 (or 2), and the final terms represent the fixed effects and the error term as specified previously. The second stage implements Equation 1 (or 2), in which the dependent variable is regressed on fitted values from the first stage regression plus the covariates. This equation is known as the structural equation, in which  $PR_{it}$  is the endogenous predictor of interest, namely, days present for student  $i$  in year  $t$ . The assumption is that the instrumental variable is uncorrelated with any omitted variables, and thus the second stage predicted value of days present will now also be uncorrelated with omitted variables as a result of implementing stage 1. In other words, the bias in the estimation of the relationship between attendance and GPA resulting from the exclusion of any omitted variables has been potentially removed with the use of the instrument.

This instrumental variables approach will be applied to the large, full sample of students and to the elementary and middle school subsamples. Before proceeding, note that it is possible that distance from school might be related to family characteristics. However, the results from Table 2 suggest otherwise. Specifically, Table 2 presents correlation coefficients and their

significance levels between distance and days present, GPA, lagged GPA, and other student and neighborhood characteristics. As mentioned previously, in the absence of parental or other family variables, neighborhood covariates often serve as proxies for such within the empirical literature (Hanushek et al., 2003). For all variables, the results indicate extremely low correlations between distance and GPA and distance and all other student and neighborhood characteristics, except for attendance. This suggests that there is nothing systematically unique about living far from school versus close in terms of its relationship to achievement, lagged achievement (which, recall, serves as a measure of student historical information), or those variables serving as proxies for family characteristics in this analysis.

Although it appears that there are no significant relationships between distance and family measures, it nonetheless is possible that families choose their residences based on distance from school.<sup>10</sup> As such, in a test of robustness, this instrumental variables approach is applied only to those students who remain at the same residence but have naturally progressed from elementary to middle school locations, in essence providing an additional exogenous source of variation in this experiment.

## Results

### Exploration of the Baseline Model

Table 3 provides parameter estimates and Huber/White/sandwich robust standard errors adjusted for classroom clustering for the model in Equation 1. The first column uses observations from the full sample of students, whereas the second and third are broken out by elementary and middle schools, respectively.<sup>11</sup> From these initial results, there are several findings related to attendance and GPA in both the full model and elementary and middle school models. First, the coefficients on days present are positive and significant in all three equations. They suggest that attending school is correlated with a higher GPA. The effect sizes, as defined by the standardized regression coefficient  $\hat{\lambda}\sigma_x/\sigma_y$ , range from  $0.24\sigma$  to  $0.34\sigma$ , thereby indicating that the attendance-achievement relationship is fairly consistent for the full sample and across elementary and middle school samples. In fact, a slightly larger coefficient in the middle school regression<sup>12</sup> demonstrates that attendance is more strongly correlated with a higher GPA as students advance through years of schooling (Roby, 2004).<sup>13</sup>

There are additional results worth noting in Table 3. First, the coefficients on student gender, race, free lunch status, English language learner, and special education are statistically significant and are consistent with much of the literature (Argys, Rees, & Brewer, 1996; Caldas & Bankston, 1997; Coates, 2003; Ogbu, 1989; Summers & Wolfe, 1977). Except for coefficients on race as “other” and special education status, the coefficients on

*Table 3*  
**Baseline Model Parameter Estimates**

	Full Sample	Elementary	Middle
Days present	0.016*** (0.000)	0.016*** (0.000)	0.020*** (0.001)
Male	-0.292*** (0.006)	-0.281*** (0.006)	-0.369*** (0.014)
Black	-0.350*** (0.011)	-0.350*** (0.012)	-0.365*** (0.024)
Hispanic	-0.236*** (0.015)	-0.238*** (0.018)	-0.230*** (0.034)
Asian	0.209*** (0.016)	0.198*** (0.016)	0.234*** (0.037)
Other	-0.247*** (0.050)	-0.254*** (0.053)	-0.247 (0.182)
Special education	-0.171*** (0.039)	-0.190*** (0.042)	-0.116 (0.074)
Free lunch recipient	-0.217*** (0.006)	-0.221*** (0.006)	-0.189*** (0.016)
English language learner	-0.200*** (0.016)	-0.215*** (0.018)	-0.126*** (0.035)
Behavior problem	-0.605*** (0.010)	-0.623*** (0.010)	-0.414*** (0.027)
Census block: percent White	0.018 (0.012)	0.025* (0.014)	-0.032 (0.035)
Census block: percent poverty	-0.188*** (0.050)	-0.159*** (0.051)	-0.371*** (0.115)
Census block: vacancy rate	-0.121*** (0.034)	-0.122*** (0.036)	-0.126*** (0.112)
Census block: log(mean income)	0.026** (0.011)	0.026** (0.011)	0.042 (0.027)
School, year, grade fixed effects	Yes	Yes	Yes
<i>n</i>	163,122	144,292	18,830
<i>R</i> <sup>2</sup>	0.19	0.19	0.22

*Note.* Huber/White/sandwich standard errors, corrected for classroom clustering, are in parentheses.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

student attributes remain statistically significant in full, elementary, and middle school regressions. Behavior problems are associated also with having lower levels of academic performance (Figlio, 2005), although the relationship appears to be slightly stronger in elementary school.

Student census-block neighborhood characteristics provide significant findings, although not consistently so: In middle school, neighborhood

characteristics seem to be less highly associated with GPA. Nonetheless, for both the elementary and middle school regressions, the coefficients indicate that a lower census block vacancy rate and lower percentage of residents at or below poverty are both associated with higher GPAs. For the elementary school sample, a higher census block percentage of Whites and higher average block income also correspond to higher student GPAs. These two relationships are not significant for the middle school sample, and this suggests that as a student progresses through years of schooling, it is the student's individual characteristics and school environment that are related to performance outcomes, not necessarily families or neighborhoods (Easton & Englehard, 1982).

### **Strengthening the Baseline Results**

In an attempt to evaluate with more precision the underlying relationship between attendance and achievement, this study next implemented a value-added model as depicted in Equation 2. In addition to all contemporaneous independent variables as described in Table 3, this model has the feature of incorporating a one-year lagged measure of student-level GPA as an independent variable, as determined by subtracting the historical model of achievement with respect to  $t - 1$  from the model with respect to  $t$ . As described, this lag is assumed to capture historical information about a student (Hanushek et al., 2003; Zimmer & Toma, 2000), and thus it serves as a proxy for an individual fixed effect.

Table 4 provides parameter estimates and Huber/White/sandwich robust standard errors adjusted for classroom clustering for the results of days present, the value-added lagged GPA component of the full model, and student characteristics. For the sake of clarity, the neighborhood covariates from Table 2, although incorporated into the model, are not presented here. To begin, the coefficients on number of days present remain positively and significantly related to GPA and consistent with the baseline model. However, incorporating the lagged measure of GPA has reduced the size of the coefficients and effect sizes of attendance in all value-added models. In conjunction with increases in  $R^2$  values for total, elementary, and middle school samples, the smaller coefficients and effect sizes indicate that the lagged value of GPA has accounted for a large portion of the total variation explained among GPA outcomes. In any case, these results provide a similar explanation as before—that there is empirical evidence that the relationship between attendance and GPA is positive and significant and that this relationship becomes slightly stronger for middle school students.<sup>14</sup>

Although adding a lag accounts for time-invariant unobservable factors predicting student achievement, it does not control for possible time-varying influences. To address this problem requires the use of the instrumental variables strategy described above. This approach is implemented for both

Table 4

Selected Parameter Estimates for Baseline Model (from Table 3) and Value-Added Model

	Full Sample		Elementary		Middle	
	Baseline	Value-Added <sup>a</sup>	Baseline	Value-Added	Baseline	Value-Added
Days present	0.016*** (0.000)	0.010*** (0.000)	0.016*** (0.000)	0.010*** (0.000)	0.020*** (0.001)	0.014*** (0.001)
One-year lagged GPA		0.512*** (0.003)		0.520*** (0.003)		0.448*** (0.009)
Male	-0.292*** (0.006)	-0.180*** (0.005)	-0.281*** (0.006)	-0.176*** (0.005)	-0.369*** (0.014)	-0.223*** (0.014)
Black	-0.350*** (0.011)	-0.204*** (0.009)	-0.350*** (0.012)	-0.209*** (0.009)	-0.365*** (0.024)	-0.209*** (0.022)
Hispanic	-0.236*** (0.015)	-0.123*** (0.011)	-0.238*** (0.018)	-0.125*** (0.013)	-0.230*** (0.034)	-0.137*** (0.033)
Asian	0.209*** (0.016)	0.123*** (0.013)	0.198*** (0.016)	0.122*** (0.014)	0.234*** (0.037)	0.124*** (0.030)
Other	-0.247*** (0.050)	-0.144*** (0.048)	-0.254*** (0.053)	-0.157*** (0.053)	-0.247 (0.182)	-0.116 (0.130)
Special education	-0.171*** (0.039)	-0.083** (0.034)	-0.190*** (0.042)	-0.091** (0.036)	-0.116 (0.074)	-0.069 (0.069)
Free lunch recipient	-0.217*** (0.006)	-0.106*** (0.004)	-0.221*** (0.006)	-0.107*** (0.005)	-0.189*** (0.016)	-0.098*** (0.014)
English language learner	-0.200*** (0.016)	-0.076*** (0.014)	-0.215*** (0.018)	-0.077*** (0.016)	-0.126*** (0.035)	-0.065** (0.029)

(continued)

**Table 4 (continued)**

	Full Sample		Elementary		Middle	
	Baseline	Value-Added <sup>a</sup>	Baseline	Value-Added	Baseline	Value-Added
Behavior problem	-0.605*** (0.010)	-0.132*** (0.009)	-0.623*** (0.010)	-0.137*** (0.010)	-0.414*** (0.027)	-0.059*** (0.025)
Neighborhood characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School, year, grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
<i>n</i>	163,122	163,122	144,292	144,292	18,830	18,830
<i>R</i> <sup>2</sup>	0.19	0.40	0.19	0.40	0.22	0.37
<i>p</i> value of Likelihood Ratio Test		<.000		<.000		<.000

*Note.* Huber/White/sandwich standard errors, corrected for classroom clustering, are in parentheses.

<sup>a</sup>As a test of robustness, the dependent variable in the value-added model was changed to  $A_{it} - A_{it(t-1)}$ . The results of those models are consistent with the value-added results here.

\*\*\* $p < .01$ .

baseline and value-added models and for full, elementary, and middle school samples. Table 5 provides results of the first stage of the instrumental variables approach. As described by Equation 3, the dependent variable is days present, and the independent variables include the instrument (i.e., the distance in exact miles a student lives from school) as well as all other covariates described in Table 1. The regressions include school, year, and grade fixed effects and Huber/White/sandwich standard errors adjusted for clustering at the classroom level. Again, for the sake of clarity, Table 5 provides only those estimates for the instrument and for student characteristics. However, neighborhood covariates are also included in the model.

The negative and statistically significant coefficients on distance indicate that as a student's mileage from school increases, the number of days that a student is present decreases. The results are consistent across models, as all coefficients on distance hover around a value of approximately  $-0.50$ . Thus, controlling for student and neighborhood characteristics, there is evidence in Table 5 indicating that student attendance tends to decrease with an increase in the distance from school in both elementary and middle school. Briefly turning to the student characteristics in the table, having a higher GPA in the previous academic year is associated with a higher number of days present in the current year. Other demographic characteristics indicate a high association with number of days present, except for free lunch recipient and behavior problem indicators. Both of these coefficients are negative and statistically significant.

Table 6 compares the coefficients of days present from the instrumental variables regressions (i.e., from the second stage of the analysis) to those from Table 4. Although both series of parameters on attendance are positive and significant, the results from the instrumental variables regressions indicate that the prediction of individual attendance on student achievement is larger in magnitude than what the ordinary least squares regressions in Table 4 had suggested.<sup>15</sup> This is evident from the coefficients in the main sample as well as in both elementary and middle school analyses. The effect sizes of the instrumental variable results on attendance range between  $0.39$  and  $0.45\sigma$  for baseline models and  $0.28$  and  $0.29\sigma$  for value-added models. This implies that a one standard deviation increase in the days a student is present in school is associated with a statistically significant  $0.39$ – $0.45$  or  $0.28$ – $0.29$  standard deviation change in GPA, and these results are larger than what Table 4 indicated. Consistent with the results in Table 4, however, the lagged achievement feature of the value-added model decreases the magnitude of the coefficients and effect sizes, thereby providing evidence that the lag accounts for a significant portion of the historical unobserved measures of current GPA.<sup>16</sup>

The results of the instrumental variables strategy indicate that distance from school provided an exogenous measure of attendance in this analysis, one that has removed potentially confounding influences of unobserved,

Table 5  
**Selected Parameter Estimates for the First Stage Instrumental Variables Approach: Baseline and Value-Added Models**

	Full Sample		Elementary		Middle	
	Baseline	Value-Added <sup>a</sup>	Baseline	Value-Added	Baseline	Value-Added
Instrument: distance from school (in miles)	-0.532*** (0.016)	-0.504*** (0.016)	-0.539*** (0.017)	-0.515*** (0.017)	-0.500*** (0.057)	-0.474*** (0.056)
One-year lagged GPA		2.143*** (0.033)		2.051*** (0.035)		2.846*** (0.099)
Male	-0.078 (0.064)	0.396*** (0.064)	-0.093 (0.069)	0.325*** (0.068)	0.028 (0.195)	1.000*** (0.195)
Black	2.379*** (0.129)	2.886*** (0.130)	1.996*** (0.143)	2.486*** (0.142)	3.995*** (0.329)	4.821*** (0.324)
Hispanic	0.561*** (0.158)	1.091*** (0.159)	0.435*** (0.175)	0.859*** (0.173)	1.687*** (0.413)	2.243*** (0.406)
Asian	7.491*** (0.204)	6.795*** (0.208)	7.058*** (0.231)	6.610*** (0.229)	8.684*** (0.522)	7.517*** (0.513)
Other	0.597 (0.741)	1.238* (0.743)	0.316 (0.812)	0.737 (0.802)	3.174 (2.014)	3.991** (1.976)
Special education	2.200*** (0.546)	2.762*** (0.542)	2.275*** (0.595)	2.588*** (0.588)	3.450*** (1.438)	3.685*** (1.411)
Free lunch eligible	-3.165*** (0.072)	-2.635*** (0.072)	-3.081*** (0.077)	-2.564*** (0.076)	-3.758*** (0.214)	-3.062*** (0.211)
English language learner	-0.274 (0.178)	0.333* (0.189)	-0.401* (0.207)	0.181 (0.205)	0.692 (0.506)	1.145** (0.496)
Behavior problem	-4.485*** (0.107)	-2.341*** (0.112)	-4.339*** (0.113)	-2.321*** (0.117)	-5.153*** (0.370)	-2.701*** (0.372)

(continued)

Table 5 (continued)

	Full Sample		Elementary		Middle	
	Baseline	Value-Added <sup>a</sup>	Baseline	Value-Added	Baseline	Value-Added
Neighborhood characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School, year, grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
<i>n</i>	163,122	163,122	144,292	144,292	18,830	18,830
<i>R</i> <sup>2</sup>	0.08	0.10	0.07	0.09	0.10	0.14

Note. Huber/White/sandwich standard errors, corrected for classroom clustering, are in parentheses.

<sup>a</sup>As a test of robustness, the dependent variable in the value-added model was changed to  $A_{it} - A_{it(t-1)}$ . The results of those models are consistent with the value-added results here.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

*Table 6*  
**Parameter Estimates of Days Present for Main Samples and the Restricted Sample of Students Transitioning From K–5 to 6–8 Schools: Instrumental Variable and Regression Results, Baseline and Value-Added Models**

	Full Sample		Elementary		Middle		Test of Validity: K–5 to 6–8 Movers	
	Baseline	Value-Added	Baseline	Value-Added	Baseline	Value-Added	Baseline	Value-Added
Instrumental variables strategy <sup>a</sup>	0.029*** (0.002)	0.019*** (0.002)	0.029*** (0.002)	0.018*** (0.002)	0.023*** (0.008)	0.016*** (0.007)	0.035*** (0.012)	0.019* (0.011)
Effect size <sup>b</sup>	0.45σ	0.29σ	0.45σ	0.28σ	0.39σ	0.28σ	0.53σ	0.28σ
Days present coefficients from Table 4	0.016*** (0.000)	0.010*** (0.000)	0.016*** (0.000)	0.010*** (0.000)	0.020*** (0.001)	0.014*** (0.001)	0.017*** (0.001)	0.012*** (0.001)
Effect size <sup>b</sup>	0.25σ	0.16σ	0.24σ	0.15σ	0.34σ	0.24σ	0.26σ	0.18σ
Hausman test for endogeneity of “days present”	31.25***	17.84***	30.10***	16.54***	28.60***	15.00***	12.97***	10.25***

*Note.* Huber/White/sandwich standard errors, corrected for classroom clustering, are in parentheses.

<sup>a</sup>All models’ post-estimation tests of significance reject the null hypothesis (at  $p < .000$ ) that the parameter on student distance is zero.

<sup>b</sup>Effect sizes are defined in this study as the standardized beta coefficient.

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ .

time-varying, student variables, such as academic motivation. For example, in the regressions from Table 4, unobserved motivation could be simultaneously affecting the measure of days present (a more motivated student will want to attend school in year  $t$ ) and GPA (a more motivated student will want to perform more highly in school in year  $t$ ). As such, from the regressions reported in Table 4, it may not be possible to separate out the effects of attendance and achievement because student motivation can be related to both independent and dependent variables. In the instrumental variables regressions, however, it has been assumed that this strategy excludes the confounding effects associated with attendance, such as motivation. Thus, when distance is used as an instrument in the regression, it minimizes the effects of unobserved, time-varying influences that simultaneously affect both dependent and independent variables.

### **Tests of Validity**

First, to test the robustness of the instrumental variables strategy, a subsample of students was assessed. In this scenario, the analysis included only those sixth-grade students who had proceeded from fifth to sixth grades and hence moved school locations from a district K–5 school to a district 6–8 school. Though they have transferred schools, the sample only included those who have continued to reside at the same home address while attending both elementary and middle schools. In other words, these students experienced a progression from one school location to the next, thereby providing this study with a natural experiment between elementary and middle school years.

It is possible that parents potentially planned their housing location preferences based on the distance from elementary school. However, with the matriculation from elementary to middle schools, there is now a new, and potentially unplanned, mileage that the student must travel from home to school, beginning in the sixth grade. There is a disruption in school distance from switching school locations, as some students now live closer to school while others may now have a new longer distance to travel.<sup>17</sup> Consequently, in this restricted sample, the distance from middle school was treated as exogenous variation (Miguel, Satyanath, & Sergenti, 2004) to examine whether previous results are robust to an alternative specification—one in which the distance from school is likely to be less sensitive to a family’s decision about housing.

A similar instrumental variables regression approach was taken on this subsample of students as was conducted in the previous set of analyses. In the first stage, days present was regressed on distance and all other covariates, including fixed effects and Huber/White/sandwich standard errors adjusted for classroom clustering. In the second stage, GPA was the outcome variable. The results are presented in Table 6 alongside the previous set of analyses.

### *Evaluating the Attendance-Achievement Relationship*

The baseline and value-added results of this regression show similar outcomes to other analyses in Table 6. That is, the value-added models have reduced the size of the coefficients from the baseline regressions, and yet, with the introduction of an instrumental variables design, the coefficients of days present remain positive and significant and are larger than under ordinary least squares regressions. Compared to the other results of the second stage of the instrumental variables regressions presented in Table 6, the coefficient in this analysis is slightly larger for the baseline model as is the standard error for the value-added model, suggesting a lower level of significance at  $p < .10$  in this second column. Both of these statistical results might have arisen because this analysis was conducted on a smaller subset of a sub-sample of students (i.e., only a subset of sixth grade students, which is itself a subset of the much larger sample). In addition, the treatment in this experience was slightly different: Here, only the effects of a potentially disruptive change in distance were evaluated. Nonetheless, the coefficients suggest, along with the other analyses presented in Table 6, that a potential causal relationship exists between attendance and GPA. That is, there appear to be positive payoffs from attending school: Students with a higher number of days present have higher GPAs.

A second test of validity examined the relationship between attendance and standardized testing outcomes, specifically SAT9 reading and math subjects. Because SAT9 scores were available for only second through fourth grades, the sample in this specification includes only third and fourth grade observations, with second grade scores used strictly as lags due to the implementation of the value-added model.<sup>18</sup>

The first section of regressions in Table 7 presents the elementary school results from Table 6 for the purposes of comparability to the preceding analyses. The second set of results examines GPA as an outcome strictly for those third and fourth grade students in the standardized testing sample. The relationship between attendance and GPA is consistent between the full elementary school sample and the standardized sample, for both the ordinary least squares regressions and instrumental variables approach as well as for both unstandardized coefficients and their effect sizes. Thus, when employing GPA as a measure of academic performance, there remains a positive and consistent association with attendance, even when examining a sub-sample of the Philadelphia elementary school population.

The interpretation of the relationship between attendance and GPA, as presented in previous analyses of this study, is also replicated by the implementation of models in which the dependent variables are standardized testing outcomes. Specifically, examining the final two sections of Table 7, the direction and significance of coefficients on attendance for SAT9 reading and math models remain consistent with those of elementary school GPA outcomes. There is a positive, statistically significant coefficient on attendance for each model and testing subject, though the table suggests a slightly

*Table 7*  
**Parameter Estimates on Days Present for Elementary School Subsample  
 for which Standardized Test Scores Were Available**

	Elementary GPA Results (Table 6)			SAT9 Sample			SAT9 Reading			SAT9 Math		
	Baseline	Value-Added	Baseline	Value-Added	Baseline	Value-Added	Baseline	Value-Added	Baseline	Value-Added	Baseline	Value-Added
Instrumental variables strategy corresponding to approach in Table 6 <sup>a</sup>	0.029*** (0.002)	0.018*** (0.002)	0.029*** (0.004)	0.018*** (0.004)	0.169** (0.077)	0.115** (0.050)	0.301*** (0.099)	0.209*** (0.067)				
Effect size <sup>b</sup>	0.45 $\sigma$	0.28 $\sigma$	0.41 $\sigma$	0.26 $\sigma$	0.16 $\sigma$	0.11 $\sigma$	0.23 $\sigma$	0.16 $\sigma$				
Days present coefficient corresponding to approach in Table 4	0.016*** (0.000)	0.010*** (0.000)	0.015*** (0.000)	0.009*** (0.000)	0.097*** (0.006)	0.054*** (0.004)	0.157*** (0.007)	0.090*** (0.006)				
Effect size <sup>b</sup>	0.24 $\sigma$	0.15 $\sigma$	0.21 $\sigma$	0.13 $\sigma$	0.10 $\sigma$	0.05 $\sigma$	0.12 $\sigma$	0.07 $\sigma$				

*Note.* Huber/White/sandwich standard errors, corrected for classroom clustering, are in parentheses. The SAT9 reading sample has  $n = 34,199$  and math has  $n = 41,019$ . These two groups are not statistically different from each other. As in Table 6, the Hausman tests for each model here reject exogeneity of the variable “days present.”

<sup>a</sup>All models’ post-estimation tests of significance reject the null hypothesis (at  $p < .000$ ) that the parameter on student distance is zero.

<sup>b</sup>Effect sizes are defined in this study as the standardized beta coefficient.

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

stronger relationship with math SAT9 performance compared to reading. This would support the particular importance of the relationship between attendance and math achievement that previous research on urban elementary school student performance has discussed (Balfanz & Byrnes, 2006).

Overall, the results throughout Table 7 suggest a similar interpretation as before—that a significant relationship exists between attendance and achievement, one that suggests causality as indicated by the instrumental variables strategy. Though the coefficients are larger and effect sizes are smaller for the standardized testing outcomes of this elementary school sample (perhaps in part due to the larger standard deviation associated with standardized test scores compared to that of GPA), the conclusion of the analyses nonetheless suggests that being present in school has implications that are generalizable to multiple performance measures.

## Conclusion

This article has contributed new evidence to the literature on the relationship between school attendance and academic outcomes, for both GPA and standardized testing performance. The relationship was evaluated for a longitudinal sample of elementary and middle school students in the Philadelphia School District over the 1994/1995 through 2000/2001 academic period. Overall, this study implemented three methodological approaches. The first began with the assessment of a baseline, contemporaneous specification of student outcomes, in which achievement was modeled on the basis of concurrent individual and neighborhood characteristics as well as school (and year and grade) fixed effects to account for unobservable time-invariant characteristics of the educational experience. The coefficients on the number of days present indicated positive, significant relationships between individual attendance and student-level achievement: Students who attend school have higher GPAs.

As a second approach to evaluating the attendance-achievement relationship, the study then introduced a lagged measure of achievement as a predictor of current achievement. Through the process by which the lagged measure was incorporated into the empirical specification, it was assumed that this covariate accounts for measured and unobservable historical factors that are associated with contemporaneous achievement. The coefficients on attendance, though generally consistent with those of the baseline model, were reduced in magnitude with the introduction of the value-added approach. This indicates that prior achievement accounts for a large portion of the historical variation in current student outcomes.

Although incorporating a lag controlled for time-invariant unobservable factors related to current student achievement (and therefore serves as a proxy for individual fixed effects), it did not account for time-varying influences. As such, this study proposed that this model alone could not separate the

relationship of attendance and achievement from unobservable, time-varying factors that might be influencing the estimates of both independent and dependent variables. Thus, this study employed a third empirical approach.

In this final method, an instrumental variables strategy was implemented in which distance was used as an instrument on attendance. The distance in miles that a student lives from school was claimed to be free from the effects of time-varying unobserved influences that simultaneously influence both attendance and achievement. The analyses, conducted on baseline and value-added models, have pointed toward evidence of causality: Controlling for student and neighborhood characteristics and school, grade, and year fixed effects, the point estimates indicated that students with higher school attendance attain higher educational outcomes in terms of both GPA and standardized testing.

The results are consistent for elementary and middle school samples, as well as in the subsequent test of validity on those students who had moved from K–5 to 6–8 schools within the district yet remained at the same home address. In these regression results, the instrumental variables coefficients are in line with the ordinary least squares regression estimates for both baseline and value-added models. This indicates that even after employing distance as an instrument of days present under these more restrictive specification requirements, the ability to empirically parse out attendance from other unobservables depicts a more accurate relationship between attendance and achievement than other, previously employed methodologies would have suggested.

The supplemental evaluation of the standardized testing outcomes of elementary school students suggests that attendance has predictive capability not only on GPA but also on reading and math subject test performance. The consistently positive and significant estimates within all three outcomes implemented in this supplemental analysis have suggested that the relationship derived between attendance and achievement can be generalizable to multiple indicators of academic success. Furthermore, because the statistical significance of the coefficients on days present is pervasive in all models and across multiple measures of achievement, the results imply that the attendance is a robust predictor of student achievement.

The focus on a single urban school district has enabled this study to document attendance as students progress through early years of their schooling experiences. The analysis in the present study has demonstrated that not only does attendance matter in a given school year, but it matters across multiple measures of achievement and it matters early on. Thus, in conjunction with several previous studies on the relationship between attendance and achievement, this article also supports the premise that a concurrent, and potentially causal, relationship exists between attendance and achievement across multiple grades in urban schools.

## *Evaluating the Attendance-Achievement Relationship*

To evaluate additional issues of attendance in urban schooling, there are three research extensions from this article. First, the data used in this study were only for elementary and middle school students. A longitudinal dataset that contains elementary, middle, and high school observations could provide insight on the relationship between early attendance and both current and future academic performance. Second, future research could entail evaluating how student attendance affects neighborhood quality via the schools they attend. With appropriate quasi-experimental techniques, it may be possible to derive causality between attendance in early years of education and achievement in high school and then link this relationship to subsequent school quality and neighborhood valuation, thereby incorporating an additional dimension to the relationship between attendance and achievement within urban environments. Finally, it should be noted that this article has focused on a single urban school district. While there are many advantages to evaluating a population of students within a single, large urban school district, it is possible that different results and interpretations may be found in other school districts of varying urbanicity. The results, thus, could be compared to those using data from additional urban districts in order to arrive at multi-district conclusions. Nonetheless, with the data and quasi-experimental methods employed in this study, this article has provided findings that there exists a statistically significant and practically meaningful relationship between attendance and achievement.

### Notes

The research here was supported in whole by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C050041 to the University of Pennsylvania. The opinions expressed are those of the author and do not represent the views of the Institute of the U.S. Department of Education. The author wishes to thank Laura Desimone, Adele E. Gottfried, Allen Gottfried, Doug Lynch, Erica Johnson, Rebecca Maynard, Todd Sinai, and Walter Theseira for advice and commentary. The author also would like to thank Robert Inman and the Philadelphia School District for providing an invaluable source of data. Finally, the author acknowledges great insight from the journal's editorial and review team, including four anonymous referees.

<sup>1</sup>The usable, analytical sample is smaller than the total sample. This is a function of having missing data on a student's record as well as the requirement of having a lagged measure of GPA for each student per year in the value-added models to follow. However, supplemental analyses suggest that the analytical sample is representative of the full sample and thus does not influence the interpretation of the results going forward.

<sup>2</sup>Teacher and classroom information is available for both elementary and middle school students. At the elementary level, teacher and classroom data represent classrooms in which students spend the entirety of the school day and year. However, for middle school students, teacher and classroom information represents homeroom information. For consistency of the proceeding analyses between the two school levels, teacher and classroom information was not utilized in the models. However, when included in an alternative specification strictly based on the elementary school sample, the results of attendance remained consistent with those presented in the article.

<sup>3</sup>To explore the potential bias for those students who stay versus leave the school district, a supplemental analysis was conducted. The sample was divided into three groups: those who stay in the district, those who started in the district but exited, and those who

did not start in the district but entered in a later grade. Iterating over many random draws of students from each group demonstrates that there are no statistical differences between the observable characteristics of the three groups.

<sup>4</sup>In more detail, Betts and Morell (1999) found that both GPA and standardized test performance were significant predictors of college academic attainment, even after controlling for student and family background and school resources.

<sup>5</sup>For the third and fourth grade sample utilized in the standardized testing regressions, GPA is also 2.4. Further, SAT9 measures, based on Normal Curve Equivalents, are 43.1 in reading and 52.2 in math. Note that as a whole, the elementary school “standardized testing” sample is statistically representative of the elementary school “GPA” sample.

<sup>6</sup>A secondary analysis of the dataset investigated whether movement in and out of receiving free lunch might influence the results. However, there is very little movement in and out of receiving free lunch per student.

<sup>7</sup>Students and families within the same school often share similar unobserved characteristics, and thus the literature suggests that school fixed effects can empirically represent family fixed effects as well (e.g., McEwan, 2003).

<sup>8</sup>Note that previous achievement is on the right-hand side of the equation. Unlike a model where the left-hand side variable is a difference between current and previous achievement, the approach utilized here does not constrain the coefficient of previous achievement to be a value of one (Todd & Wolpin, 2003).

<sup>9</sup>In an alternative specification, a model that incorporated student fixed effects (by coding dummy variables for each student ID number) was tested. The effects of attendance—the parameter of key interest—were quantitatively similar to the coefficients using the value-added model. As such, the original empirical specification is executed from this point forward, though the results of this particular analysis are available upon request.

<sup>10</sup>Because the Philadelphia public schools in this sample during the time-frame of this analysis are not schools of choice, it is only possible for families to “choose” schools in essence by selecting residential locations. On the other hand, parents and students cannot choose which public school to attend once living at their residential location.

<sup>11</sup>In an alternative specification, the assumption of a constant treatment of grade level was tested within both elementary and middle school samples. However, the results provide evidence of similar outcomes of attendance for grades within elementary school and within middle school.

<sup>12</sup>A Wald test was employed to assess if significant differences exist between elementary and middle school regression coefficients. The result of this test has indicated that the coefficients do in fact differ across schools. Thus, having separate regression models for elementary and middle school samples is appropriate.

<sup>13</sup>A secondary test examined the relationship between year fixed effects and attendance in order to determine if attendance in previous years can influence attendance in later years. However, a test of joint significance rejects the hypothesis of having separate, interacted coefficients between attendance and calendar year for each student.

<sup>14</sup>As before, a Wald test indicates significant statistical differences between attendance coefficients on elementary and middle school regressions.

<sup>15</sup>The results of a Hausman test of endogeneity for each instrumental variables model also suggest that, statistically, this empirical strategy is a more appropriate method for describing the relationship between days present and GPA.

<sup>16</sup>As a test of robustness, measures of attendance from year  $t - 1$  were included in the value-added model. However, the coefficients were non-significant predictors of achievement. This is consistent with the assumption that the lagged GPA term accounts for historical student information. When the lagged measure of absence is included in the baseline models, then lagged attendance becomes significant and positive (though the size of the coefficient is less than one-quarter of the size of the current attendance parameter). However, for consistency between baseline and value-added models, lagged attendance was not included in the specifications in the main analyses.

<sup>17</sup>Note that, like in Table 2, there are no highly correlated relationships except for distance and attendance.

<sup>18</sup>The third and fourth grade sample is statistically representative of the entire elementary school sample used in previous analyses within this study.

## References

- Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*, 87–107.
- Anaya, G. (1999). College impact on student learning: Comparing the use of self-reported gains, standardized test scores, and college grades. *Research in Higher Education, 40*, 499–526.
- Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management, 15*, 623–645.
- Balfanz, R., & Byrnes, V. (2006). Closing the mathematics achievement gap in high poverty middle schools: Enablers and constraints. *Journal of Education for Students Placed at Risk, 11*, 143–159.
- Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis*. Baltimore, MD: Center for Research on the Education of Students Placed at Risk, Center for Social Organization of Schools, Johns Hopkins University.
- Barrington, B. L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *Journal of Educational Research, 82*, 309–319.
- Betts, J. R., & Morell, D. (1999). The determinants of undergraduate grade point average: The relative importance of family background, high school resources, and peer group effects. *The Journal of Human Resources, 34*, 268–293.
- Borland, M. V., & Howsen, R. M. (1998). Effect of student attendance on performance: Comment on Lamdin. *Journal of Educational Research, 91*, 195–197.
- Broadhurst, K., Patron, K., & May-Chahal, C. (2005). Children missing from school systems: Exploring divergent patterns of disengagement in the narrative accounts of parents, careers, children, and young people. *British Journal of Sociology of Education, 26*, 105–119.
- Caldas, S. J. (1993). Reexamination of input and process factor effects in public school achievement. *Journal of Educational Research, 86*, 206–214.
- Caldas, S. J., & Bankston, C., III. (1997). Effect of school population socioeconomic status on individual academic achievement. *Journal of Educational Research, 90*, 269–277.
- Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian, and East Asian high school students. *Child Development, 66*, 1215–1234.
- Coates, D. (2003). Education production functions using instructional time as an input. *Education Economics, 11*, 273–292.
- Connell, J. P., Spencer, M. B., & Aber, J. L. (1994). Educational risk and resilience in African-American youth: Context, self, action, and outcomes in school. *Child Development, 65*, 493–506.
- Corville-Smith, J. (1995). Truancy, family processes, and interventions. In B. Ryan, G. Adams, T. Gullotta, R. Weissberg, & R. Hampton (Eds.), *The family-school connection: Theory, research, and practice* (pp. 270–287). Thousand Oaks, CA: Sage.
- Coutts, D. (1998). How to better track effective school indicators: The control chart techniques. *American Secondary Education, 27*, 2–10.
- Devadoss, S., & Foltz, J. (1996). Evaluation of factors influencing student class attendance and performance. *American Journal of Agriculture Economics, 78*, 499–507.
- Dreyfoos, J. (1990). *Adolescents at risk: Prevalence and prevention*. New York: Oxford University Press.
- Durden, G. C., & Ellis, L. V. (1995). The effects of attendance on student learning in principles of economics. *American Economic Review, 85*, 343–346.

- Easton, J., & Englehard, G., Jr. (1982). A longitudinal record of elementary school absence and its relationship to reading achievement. *Journal of Educational Research*, 75, 269–274.
- Ekstrom, R. B., Goertz, M. E., Pollack, J. M., & Rock, D. A. (1986). Who drops out of high school and why? Findings from a national study. *Teachers College Record*, 87, 356–373.
- Epstein, J. L., & Sheldon, S. B. (2002). Present and accounted for: Improving student attendance through family and community involvement. *Journal of Educational Research*, 95, 308–318.
- Figlio, D. N. (2005). *Boys named Sue: Disruptive children and their peers* (NBER Working Paper No. 11277). Cambridge, MA: NBER.
- Fine, M. (1991). *Framing dropouts: Notes on the politics of an urban public high school*. Albany: State University of New York Press.
- Fine, M. (1994). Chartering urban school reform. In M. Fine (Ed.), *Chartering urban school reform* (pp. 5–30). New York: Teachers College Press.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, 59, 117–142.
- Finn, J. D. (1993). *School engagement and students at risk*. Washington, DC: National Center for Education Statistics.
- Frenette, M. (2004). Access to college and university: Does distance to school matter? *Canadian Public Policy*, 30, 427–443.
- Halfon, D., Vevea, J. L., Iritani, B., Cho, H., Khatapoush, S., & Saxe, L. (2002). Truancy, grade point average, and sexual activity: A meta-analysis of risk indicators for youth substance use. *Journal of Social Health*, 72, 205–211.
- Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18, 527–544.
- Jensen, P., & Nielsen, H. S. (1997). Child labour or school attendance? Evidence from Zambia. *Journal of Population Economics*, 10, 407–424.
- Johnson, G. M. (2005). Student alienation, academic achievement, and WebCT use. *Educational Technology and Society*, 8, 179–189.
- Kane, J. (2006). School exclusions and masculine, working-class identities. *Gender and Education*, 18, 673–685.
- King, A. R. (2000). Relationships between CATI personality disorder variables and measures of academic performance. *Personality and Individual Differences*, 29, 177–190.
- Kirby, A., & McElroy, B. (2003). The effect of attendance on grade for first year economics students in University College Cork. *The Economic and Social Review*, 34, 311–326.
- Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75, 63–82.
- Lamdin, D. J. (1996). Evidence of student attendance as an independent variable in education production functions. *Journal of Educational Research*, 89, 155–162.
- Lehr, C. A., Sinclair, M. F., & Christenson, S. L. (2004). Addressing student engagement and truancy prevention during the elementary school years: A replication study of the Check & Connect model. *Journal of Education for Students Placed at Risk*, 9, 279–301.
- Levine, J. R. (1992). *The effect of different attendance policies on student attendance and achievement*. Paper presented at the annual meeting of the Eastern Psychological Association, Boston, MA. (ERIC Document Reproduction Service No. ED 348 762)

## *Evaluating the Attendance-Achievement Relationship*

- Marburger, D. R. (2001). Absenteeism and undergraduate exam performance. *Journal of Economic Education, 14*, 11–16.
- Marburger, D. R. (2006). Does mandatory attendance improve student performance? *Journal of Economic Education, 37*, 148–155.
- McEwan, P. J. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review, 22*, 131–141.
- Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy, 112*, 725–753.
- Miller, R. T., Murnane, R. J., & Willett, J. B. (2008). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis, 30*, 181–200.
- Neild, R. C., & Balfanz, R. (2006). An extreme degree of difficulty: The educational demographics of urban neighborhood schools. *Journal of Education for Students Placed at Risk, 11*, 123–141.
- Newmann, F. (1981). Reducing student alienation in high schools: Implications of theory. *Harvard Educational Review, 51*, 546–564.
- Nichols, J. D. (2003). Prediction indicators for students failing the state of Indiana high school graduation exam. *Preventing School Failure, 47*, 112–120.
- Ogbu, J. U. (1989). The individual in collective adaptation: A framework for focusing on academic underperformance and dropping out among involuntary minorities. In L. Weis, E. Farrar, & H. G. Petrie (Eds.), *Dropouts from school: Issues, dilemmas, and solutions*. (pp. 181–204). Albany: State University of New York Press.
- Orfield, G., & Kornhaber, M. L. (Eds.). (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: Century Foundation Press.
- Orfield, G., Losen, D., Wald, J., & Swanson, C. B. (2004). *How minority youth are being left behind by the graduation rate crisis*. Washington, DC: The Urban Institute.
- Phillips, M. (1997). What makes schools effective? A comparison of the relationships of communitarian climate and academic climate to mathematics achievement and attendance during middle school. *American Education Research Journal, 34*, 633–662.
- Roby, D. E. (2004). Research on school attendance and student achievement: A study of Ohio schools. *Educational Research Quarterly, 28*, 3–14.
- Romer, D. (1993). Do students go to class? Should they? *Journal of Economic Perspectives, 7*, 167–174.
- Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Education Research Journal, 32*, 583–625.
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Education Research Journal, 42*, 3–42.
- Rumberger, R. W., & Thomas, S. L. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education, 73*, 39–67.
- Schlossberg, M., Greene, J., Phillips, P. P., Johnson, B., & Parker, B. (2006). School trips: Effects of urban form and distance on travel mode. *Journal of the American Planning Association, 3*, 337–346.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Education Research Association.
- Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics, 74*, 199–250.

- Sheldon, S. B. (2007). Improving student attendance with school, family, and community partnerships. *Journal of Educational Research, 100*, 267–275.
- Stouthamer-Loeber, M., & Loeber, R. (1988). The use of prediction data in understanding delinquency. *Behavioral Sciences and the Law, 6*, 333–354.
- Summers, A. A., & Wolfe, B. I. (1977). Do schools make a difference? *American Economic Review, 67*, 639–652.
- Swanson, C. (2004). *Who graduates? Who doesn't? A statistical portrait of public high school graduation, class of 2001*. Washington, DC: The Urban Institute.
- Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal, 113*, F3–F33.
- Van Matre, J. C., Valentine, J. C., & Cooper, H. (2000). Effects of students' after-school activities on teachers' academic expectations. *Contemporary Educational Psychology, 25*, 167–183.
- Wang, X., Blomberg, T. G., & Li, S. D. (2005). Comparison of the educational deficiencies of delinquent and nondelinquent students. *Evaluation Review, 29*, 291–312.
- Wasley, P. A. (2002). Smaller classes, small schools etc. *Educational Leadership, 59*, 6–10.
- Willingham, W. W., & Breland, H. M. (1982). *Personal qualities and college admissions*. Princeton, NJ: The College Board.
- Zimmer, R. W., & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management, 19*, 75–92.

Manuscript received November 12, 2008

Final revision received August 26, 2009

Accepted August 31, 2009