

Examining Heterogeneity in the Effect of Taking Algebra in Eighth Grade

JORDAN H. RICKLES

University of California, Los Angeles

ABSTRACT. Increased access to algebra was a focal point of the National Mathematics Advisory Panel's 2008 report on improving mathematics learning in the United States. Past research found positive effects for early access to algebra, but the focus on average effects may mask important variation across student subgroups. The author addresses whether these positive effects hold up when the analysis is expanded to examine effect heterogeneity. Using a nationally representative sample of eighth-grade students in 1988, the author examined sensitivity of findings to methods for selection bias adjustment, heterogeneity across the propensity to take algebra in Grade 8, and across schools. The findings support past research regarding positive benefits to Grade 8 algebra and are consistent with policies that increase access to algebra in middle school.

Keywords: algebra, causal inference, course taking, propensity score analysis

In most developed nations, all students are expected to study algebra and geometry in the middle grades (Schmidt, 2004). In the United States, however, most students take their first formal algebra class in high school, with just over a third of students taking algebra or a more advanced mathematics course in Grade 8 (U.S. Department of Education, 2009). Given the sequential structure of mathematics course taking in high school, and to a lesser extent science, algebra acts as a gatekeeper for access to advanced courses and college-readiness. As such, a push for greater access to algebra in Grade 8 has grown since the mid-1990s and was a key component of the National Mathematics Advisory Panel's 2008 recommendations for improving mathematics instruction in the United States (National Mathematics Advisory Panel, 2008). As the federal government and states continue to make strides toward a common curriculum, the question about when students should take their first formal algebra class becomes increasingly important.

The rationale for algebra in Grade 8 stems from the hypothesis that early access gives students a better chance to succeed in high school mathematics and take the advanced mathematics courses necessary for college admission. Opponents of Grade 8 algebra claim unprepared students will face mathematics frustration and failure that could push students away from future mathematics courses and possibly drop out

of school. The selective nature of algebra placement carried out in most districts across the country means schools can limit Grade 8 algebra to the highest achieving students at the possible cost of restricting opportunities for some, or take a more egalitarian, or universal, approach at the possible cost of misaligning instruction for under-prepared and advanced students. While the research literature is not silent on this debate, neither does it provide a definitive answer to the question of whether a Grade 8 student is better off in an algebra or prealgebra classroom. Previous research on the effects of algebra under a selective placement process typically find positive benefits for earlier access to algebra (Stein, Kaufman, Sherman, & Hillen, 2011), but the findings speak to the more general question of whether the average student benefits from algebra. If different types of students experience different effects from algebra, it is not clear how well the effects for an average student apply to all students. Gamoran and Hannigan (2000) found, for example, that low-achieving students who only take algebra in Grade 8 do no better, or maybe worse, than those who never take algebra. This finding of a differential effect raises questions about the generalizability of other studies that fail to examine effect heterogeneity. Of particular importance to the current policy debate is whether those not taking algebra in Grade 8 would benefit from a more egalitarian placement process. If heterogeneous effects exist, inferences about the benefits of Grade 8 algebra for the average student, or for the types of students who typically take algebra in Grade 8, may not apply to students targeted by curriculum changes.

Given the selective nature of course placement, with Grade 8 algebra typically reserved for students who demonstrate high mathematics aptitude, findings from observational or quasi-experimental studies potentially suffer from selection bias. As such, the extent to which it is possible to infer a causal relationship between course placement and a given outcome depends on the confidence that the study design adequately accounts for the differential factors related to placement and the outcomes. For example, Smith (1996)

Address correspondence to Jordan H. Rickles, 300 Charles E. Young Drive North, GSEIS Building, 3rd Floor, University of California, Los Angeles, Los Angeles, CA 90095-1522, USA. (E-mail: jruckles@ucla.edu)

controlled for a limited set of student characteristics within a path analysis while Gamoran and Hannigan (2000) included a richer set of control variables in a regression model to account for preexisting differences in students who took algebra and those who did not.

One increasingly popular method to account for selection bias is to directly model the selection process and adjust your sample based on the propensity for treatment selection (Rosenbaum & Rubin, 1983; Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007; Morgan & Winship, 2007). This approach was employed by Attewell and Domina (2008), for example, to analyze the effects of curricular intensity on academic performance, but has not been used to directly examine the effects of Grade 8 algebra. A propensity score approach explicitly places the analysis within the potential outcomes framework (Holland, 1986; Rubin, 1974) and, under certain assumptions, can result in unbiased estimates (Rosenbaum & Rubin, 1983, 1984).

In this article, I apply the potential outcomes framework for causal inference and the propensity score method to formally examine whether Grade 8 algebra effects differ across students and schools to shed light on the possible gap in a literature focused on the average student. Using a nationally representative sample of eighth-grade students in 1988, I look at whether taking algebra in Grade 8 affects high school mathematics achievement. Following a diagnostic routine developed by Morgan and Todd (2008), I first examine whether traditional regression-based estimates of course-taking effects fail to identify heterogeneous effects between students who took algebra and those who did not take algebra. Second, I investigate effect heterogeneity across students with different propensities for taking Grade 8 algebra. Last, I explore the extent to which effect heterogeneity exists across schools. Before presenting the results, I provide background on the potential outcomes framework and prior research related to taking algebra in Grade 8. I conclude the article by discussing the limitations and robustness of the findings.

The Potential Outcomes Framework, Selection Bias, and Effect Heterogeneity

Publicly and politically, the push to expand algebra to more eighth-grade students stems from the basic observation that eighth-grade students in algebra exhibit better mathematics performance than eighth-grade students not in algebra. Yet the jump from an observed, cross-sectional difference between groups to a causal inference about the effect of algebra assumes equivalent groups and results in what Holland (1986) called the “prima facie causal effect” (p. 949) and what Morgan and Harding (2006) called the “naïve estimator of the average causal effect” (p. 10). The quality of this naive claim depends on how students select, or are placed, in algebra. If the selection process systematically favors higher performing students, or students with characteristics associated with greater academic suc-

cess, there is a classic case of selection bias (Shadish, Cook, & Campbell, 2002). It is not possible to determine whether observed differences in an outcome (Y) are the result of taking algebra ($Z = 1$) or not ($Z = 0$), or the result of pre-existing characteristics (X) associated with both Z and Y .

The potential outcomes framework for causal inference popularized by Rubin (1974) formalizes the problem of identifying the effect of Grade 8 algebra on mathematics performance. Under this framework, an effect is defined at the unit of analysis based on each unit’s potential outcomes under exposure and no exposure. For the current case, the effect of Grade 8 algebra (δ) for individual i is defined as

$$\delta_i = y_t(i) - y_c(i), \quad (1)$$

where $y_t(i)$ is the outcome for individual i if she or he takes algebra in Grade 8 (treatment) and $y_c(i)$ is the outcome for individual i if she or he does not take algebra in Grade 8 (control). For a given population, the population average treatment effect is defined as

$$E[\delta] = E[Y_t] - E[Y_c], \quad (2)$$

where $E[Y_t]$ is the average outcome for all individuals if they took algebra in Grade 8 and $E[Y_c]$ is the average outcome for all individuals if they did not take algebra in Grade 8. For a more detailed discussion of the potential outcomes framework see Holland (1986), Morgan and Winship (2007), and Schneider et al. (2007).

The fundamental problem of causal inference (Holland, 1986) is that y_t and y_c are never observed for the same individual. In reality, it is only possible to observe y_t for students who took algebra ($Z = 1$) and y_c for students who did not ($Z = 0$). As a result, it is necessary to fall back on estimating average group effects under certain assumptions, such as selection independence and unit homogeneity (Holland, 1986), where $E(Y_t | Z = 0) = E(Y_t | Z = 1)$ and $E(Y_c | Z = 1) = E(Y_c | Z = 0)$. Thus, given these assumptions, it is possible to estimate an average treatment effect (ATE) based on observed outcomes such that

$$E[\delta_{ATE}] = E[Y_t | Z = 1] - E[Y_c | Z = 0]. \quad (3)$$

Winship and Morgan (1999) laid out how the degree to which the ATE defined by Equation 3 represents the true average treatment effect depends on “the way in which individuals are assigned (or assign themselves) to the treatment and control groups” (p. 665). This assignment can result in two sources of bias that parallel the assumptions of independence and homogeneity (Holland, 1986).

The first source of bias, which is conventionally described as selection bias (Shadish et al., 2002), is from pre-existing differences in the treatment and control groups so that one group would do better than the other regardless of treatment receipt. Given the hypothesized selection and placement in Grade 8 algebra discussed previously, selection bias is all but

a given in this case because, at the very least, students with a history of high mathematics achievement are more likely to take Grade 8 algebra and more likely to exhibit higher mathematics achievement in the future relative to students who do not take Grade 8 algebra.

The second, less obvious, source of potential bias is the heterogeneity in the treatment effect between the treatment and control groups. This heterogeneity can be assessed by estimating the average treatment effect on the treated (ATT) and the average treatment effect on the control (ATC). Based on individual potential outcomes, the ATT and ATC can be defined as

$$ATT : E[\delta_{ATT}] = E[Y_t|Z = 1] - E[Y_c|Z = 1]; \quad (4a)$$

$$ATC : E[\delta_{ATC}] = E[Y_t|Z = 0] - E[Y_c|Z = 0]. \quad (4b)$$

Under the rival positive and negative selection hypotheses (discussed in the next section), heterogeneity in the ATT and ATC would be expected. If the positive selection hypothesis holds, the ATT should be greater than the ATC (i.e., those placed in Grade 8 algebra are more likely to benefit from Grade 8 algebra than those not placed in the course). If the negative selection hypothesis holds, the ATC should be greater than the ATT (i.e., those excluded from Grade 8 algebra are more likely to benefit from Grade 8 algebra than those placed in the course). Additionally, a policy directive that expands algebra to eighth-grade students not currently taking algebra hinges on the assumption that the ATT and ATC are positive. Thus, for policy analysis it is important to estimate the ATT and ATC, or at least examine effect heterogeneity across subgroups associated with treatment assignment.

Studies that seek to measure the effect of educational practices like tracking, ability grouping, and curricular intensity typically employ a regression modeling technique to adjust for the first source of bias (i.e., selection bias) by including a vector of potential confounding factors (\mathbf{X}) in a model that estimates the effect of Z on Y . The simple ordinary least squares (OLS) regression model would take the following form:

$$\hat{Y}_i = \alpha + \delta Z_i + \beta \mathbf{X}_i, \quad (5)$$

where the estimated treatment effect (δ) is conditional on the vector of covariates (\mathbf{X}). What is missing, however, and hard to conceptualize within the regression approach, is an estimate of the ATT and ATC to measure the second source of bias.

Another approach to estimate causal effects is to match the treatment and control groups so that treatment receipt is independent of other factors (Ho, Imai, King, & Stuart, 2007; Morgan & Harding, 2006; Schafer & Kang, 2008; Stuart & Rubin, 2008). In essence, this is what random assignment does. Under random assignment, a random selection mechanism is employed to assign units to a treatment

or control group. This random mechanism makes receipt of Z independent of \mathbf{X} , and thus estimates of $E(Y|Z)$ will be unbiased. In many social science research situations where the selection mechanism(s) is not random, one can seek to equate treatment and control groups based on the estimated propensity score (Rosenbaum & Rubin, 1983). A propensity score is the predicted probability that a unit selected (or was placed) in the treatment group ($Z = 1$) versus the control group ($Z = 0$) given a set of observed covariates (\mathbf{X}), most typically estimated using a logit or probit model. For example, using the same notation for Equation 5, the logistic regression model to generate predicted probabilities (i.e., propensity scores) would be

$$Pr(Z_i = 1) = p_i = \exp^{\alpha + \beta \mathbf{X}_i} / (1 + \exp^{\alpha + \beta \mathbf{X}_i}). \quad (6)$$

The propensity score can then be used in a variety of ways for a nonparametric correction of the treatment and control group differences across \mathbf{X} . Popular uses of the propensity score include direct matching, subclassification, and weighting (Ho et al., 2007; Morgan & Harding, 2006; Schafer & Kang, 2008).

Morgan and Harding (2006) provided a detailed discussion of propensity score methods for estimating causal effects, including the benefits of the propensity score approach versus a regression-based approach. These benefits include a more explicit focus on heterogeneity of causal effects (i.e., more explicit focus on whether one is estimating the ATE, ATT, or ATC), more explicit focus on the distribution and overlap of covariates between the treatment and control groups, and reduced concerns about the functional form of a statistical model to estimate causal effects. Ho et al. (2007) and Schafer and Kang (2008) also discussed the double robustness benefit of combining regression-based methods and propensity score methods.

It is important to understand, however, that like the regression-based methods, the propensity score methods cannot control for selection bias due to unobserved factors. While random assignment creates treatment and control groups that should be equivalent along all observed (\mathbf{X}) and unobserved (U) preconditions, the propensity score approach must invoke a conditional independence assumption often referred to as strong ignorability (Holland, 1986; Rosenbaum & Rubin, 1983). That is, it must be assumed the probability of selection conditioned on the observed covariates is the same as the probability of selection conditioned on the observed and unobserved covariates, or that $E(Z|\mathbf{X}) = E(Z|\mathbf{X}, U)$.

Another complication when estimating effects of educational practices like tracking, ability grouping, and curricular intensity is that what constitutes the actual treatment may vary across schools. Under the potential outcomes framework, variation in the treatment condition across schools can result in a break of the stable-unit-treatment-value assumption (SUTVA; Rubin, 1978, 1980), or what Hong and Raudenbush (2006) referred to as treatment enactment

variation due to an organizational effect. This variation can result in average treatment effect heterogeneity across schools. One can, however, investigate school-level effect heterogeneity through multilevel modeling (Gitelman, 2005) to better understand how well an overall average treatment effect estimate will generalize to students in different schooling environments.

Placement in Grade 8 Algebra

As mathematics content, algebra is seen as generalized arithmetic, the mastery of which requires that students first learn arithmetic operations with whole numbers, fractions, and decimals (Wu, 2001). With the primary schooling grades emphasizing the foundational mathematics content, students are typically not prepared to master formal algebra until Grade 7 or Grade 8. Gaps or breakdowns in mathematics instruction during the primary grades may require prealgebra instruction during the middle grades and delay a student's algebra readiness until high school. In addition to mathematics preparation, a theoretical and empirical discussion of how students select, or are placed, in an algebra or prealgebra course when they enter Grade 8 is closely connected with any discussion regarding tracking and ability group placement. In mathematics, Grade 8 algebra could be an entry point for a college-prep high school track or the consequence of ability groupings in earlier grades. Thus, to better understand what factors are associated with the Grade 8 algebra selection process it is possible to borrow from a broader body of research.

Oakes and Guiton (1995) reviewed different theories about tracking and how schools decide to place students in different courses. The predominant theory stems from the rational choice or human capital theories in economics, in which schools try to match students and courses in a way that effectively accommodates differences in student abilities. Other theories, however, point to cultural, organizational, and political constraints to the efficiency model. Combined with findings from their own research, Oakes and Guiton conclude that

tracking decisions result from the synergy of three powerful factors: differentiated, hierarchical curriculum structures; school cultures alternatively committed to common schooling and accommodating differences; and political actions by individuals within those structures and cultures aimed at influencing the distribution of advantage. (p. 30)

Similar research supports a notion of selection based on a rational choice process that is moderated and constrained by many factors, including social class, teacher and parent input, student expectations, and school assignment practices (Kelly, 2004; Smith, 1996; Spade, Columba, & Vanfossen, 1997; Useem, 1992).

Similarly, in their work on college selection, Brand and Xie (2010) described two selection hypotheses. The first is akin to the economics-based rational choice model in which those most likely to participate in college (or in this

case Grade 8 algebra) are the ones most likely to benefit from participation. They called this the positive selection hypothesis. The alternative hypothesis has its roots in sociology and is more in line with theories about cultural, organizational, and political constraints described by Oakes and Guiton (1995). Under this model individuals who are least likely to participate in college (or, again, in this case Grade 8 algebra) are the most likely to benefit from participation. Brand and Xie called this the negative selection hypothesis and their research on college selection supports this hypothesis.

The competing theories for college selection and high school tracking placement can also apply to selection in Grade 8 algebra. By examining effect heterogeneity, this article tests whether the positive or negative selection hypothesis best explains Grade 8 algebra selection. If the positive selection hypothesis holds, one should expect students with a high likelihood for Grade 8 algebra selection to benefit more than those with a low likelihood. Conversely, if the negative selection hypothesis holds, one should expect students with a low likelihood for Grade 8 algebra selection to benefit more than those with a high likelihood.

Prior Studies on Grade 8 Algebra

Most research related to taking algebra in Grade 8 addresses the broader issues of tracking, ability grouping, and course-taking patterns, and not the specific issue of Grade 8 algebra. In a review of the research on ability grouping, Slavin (1990) found no overall difference in academic achievement between students grouped by ability and those in heterogeneous groupings. Slavin's conclusion has been challenged (Hallinan, 1990), however, and many studies find a positive association between the number of mathematics courses taken in high school and future academic achievement (Attewell & Domina, 2008; Lee, Croninger, & Smith, 1997). In their review of studies that focus specifically on algebra, Stein et al. (2011) found that research based on selective early algebra course taking suggests overall positive effects on academic achievement, while the results from research based on universal algebra policies are more mixed.

One of the earlier studies that aimed to estimate the effect of Grade 8 algebra was conducted by Smith (1996). Using High School and Beyond data for the 1980 cohort of 10th-grade students, Smith examined the effect of early access to algebra—defined as taking algebra in Grade 8—on mathematics attainment 2 years later (Grade 12) and concluded that “early access to algebra has a sustained positive effect on students, leading to more exposure to advanced mathematics curriculum and, in turn, higher mathematics performance by the end of high school” (p. 148). Smith used a path analysis design that controlled for student ethnicity, gender, and socioeconomic status (SES) to account for selection bias. The path analysis also accounted for Grade 10 mathematics performance, whether they were on an academic track or not, and educational aspirations in the estimate of

Grade 12 achievement. This analysis, however, has two main limitations. First, with a relatively narrow set of covariates to account for all the factors likely to influence enrollment in Grade 8 algebra and mathematics performance, the assumption of strong ignorability is not likely to hold, and selection bias probably still exists. Second, restricting the analysis of a Grade 8 treatment to data collected on a Grade 10 cohort likely biases the results by conditioning on posttreatment factors. Furthermore, the path analysis does not facilitate investigation of treatment effect heterogeneity.

The study conducted by Gamoran and Hannigan (2000) looked more broadly at the effect of taking algebra in high school but also provided estimates of the effect for students who took algebra in Grade 8. They used National Education Longitudinal Study (NELS) data for a 1988 cohort of eighth-grade students to examine the effect of taking algebra on Grade 10 mathematics performance. The authors concluded that all students benefit from taking algebra, but also noted that low-achieving students who only take algebra in Grade 8 do no better, or maybe worse, than those who never take algebra. An OLS regression model was used to estimate the effect of taking algebra controlling for student demographics (gender, ethnicity, SES, limited English proficiency) and Grade 8 test performance in mathematics, reading, science, and history. Interactions between algebra enrollment and Grade 8 mathematics performance were included to investigate effect heterogeneity across student mathematics ability. While Gamoran and Hannigan conditioned on more confounding factors than Smith, both studies relied on regression modeling to estimate a causal effect and thus their conclusions hinge on the modeling assumptions and do not provide a clear definition of the ATE, ATT, or ATC.

Ma (2005) took a slightly different approach, employing hierarchical linear growth modeling to examine the effect of early access to algebra—defined as algebra in middle school—on growth in four mathematics areas (basic skills, algebra, geometry, and quantitative literacy). Using Longitudinal Study of American Youth data for a 1987 cohort of seventh-grade students who progressed to Grade 12, Ma found that low-achieving middle school students who took algebra exhibited mathematics performance growth that was “not only faster than low achieving students who were not accelerated into formal algebra but also faster than high achieving students who were not accelerated into formal algebra” (pp. 452–453). This finding supports the negative selection hypothesis but contradicts the findings reported by Gamoran and Hannigan (2000). A major benefit to growth modeling is that any time-invariant factors, observed or unobserved, are controlled. Ma also included a series of student and school characteristics in the model to further restrict potential bias.

Studies that looked at universal policies typically used an interrupted time-series research design to identify the effect of algebra. For example, Burris, Heubert, and Levin (2006) studied students in a suburban New York school district during the 1990s to examine effects of taking accelerated

mathematics—primarily defined as taking algebra in Grade 8. The analysis followed six student cohorts from middle school through high school, where the first three cohorts were not exposed to the district-mandated accelerated mathematics policy and the other three cohorts were exposed. The researchers found that the probability of completing advanced mathematics courses increased significantly for all groups exposed to accelerated mathematics, and creating heterogeneous mathematics classes did not negatively affect initial high achievers. Allensworth, Nomi, Montgomery, and Lee (2009) used a similar approach to examine the effects of a mandatory college-prep curriculum, where a key component was Grade 9 algebra, for Chicago high school students. They found few benefits from the mandatory curriculum. The time series approach helps mitigate selection bias, but raises concerns about history bias (Shadish et al., 2002).

Reviewing the research literature on the effects of algebra reveals how researchers employed different research designs to estimate causal effects. Yet besides studies of universal policy changes, they primarily relied on regression modeling techniques to account for potential selection bias, and none of the studies explicitly examined causal effects within the potential outcomes framework to parse out heterogeneous effects. With this article, I present evidence to address this gap in the literature with regards to Grade 8 access to algebra under more general, selective course assignment processes.

Method

The overall purpose of this study was to test the homogeneity of findings regarding the effect of Grade 8 algebra when placed within a potential outcomes framework. More specifically, I tested for heterogeneity in three ways: (a) comparing traditional regression-based estimates to inverse propensity score weighted regression estimates of the ATT and ATC, (b) comparing effect estimates across propensity score subclassification, and (c) examining school-level variance in the effect estimates.

Regression Models

Given the traditional emphasis on regression-based methods for controlling selection bias, I first estimated the effect of Grade 8 algebra using an OLS regression model in which the dependent variable (Y) was predicted from a vector (\mathbf{X}) of student characteristics, family characteristics, student academic history, and school-level characteristics, and a dichotomous variable (Z) indicating participation in Grade 8 algebra. This model follows the form of Equation 5.

In the Results section I refer to estimated effects from three different models. The first model (Model 1) is the naive estimate model that excludes \mathbf{X} . The second model (Model 2) uses a restricted set of covariates in \mathbf{X} to parallel the analysis conducted by Smith (1996). The covariates in the restricted set are gender, ethnicity, and family SES. The

third model (Model 3) includes the full set of covariates discussed subsequently.

Inverse Propensity Score Weighting

Following Morgan and Todd (2008), I compared the OLS regression-based effect of algebra to weighted regression-based effects, where the weights were constructed from estimated propensity scores. I estimated the propensity of taking algebra in Grade 8 using a logistic regression model based on a vector (\mathbf{X}) of student characteristics, family characteristics, student academic history, and school-level characteristics (see Equation 6). To estimate the ATT, the weights take the following form:

$$\begin{aligned} \text{For } z_i = 1 : w_i &= 1; \\ \text{For } z_i = 0 : w_i &= p_i / (1 - p_i). \end{aligned} \quad (7)$$

To estimate the ATC, the weights take the following form:

$$\begin{aligned} \text{For } z_i = 1 : w_i &= (1 - p_i) / p_i; \\ \text{For } z_i = 0 : w_i &= 1. \end{aligned} \quad (8)$$

I also estimated the ATE with the following weights:

$$\begin{aligned} \text{For } z_i = 1 : w_i &= 1/p_i; \\ \text{For } z_i = 0 : w_i &= 1/(1 - p_i). \end{aligned} \quad (9)$$

If the effects of algebra are constant across individuals, or at least vary randomly with respect to selection, then, the ATT and ATC estimates should not differ. If the ATT estimate is greater than the ATC estimate, then, it can be concluded that algebra has a larger estimated effect for those who already take algebra than those who do not take algebra. This finding would be consistent with the positive selection hypothesis. Conversely, if the ATC estimate exceeds the ATT estimate, then taking algebra has a larger estimated effect for those who do not take algebra. This finding would be consistent with the negative selection hypothesis.

Propensity Score Subclassification

For the second approach to examine effect heterogeneity, I looked at variation in the estimated effects across propensity score subclasses (Rosenbaum & Rubin, 1984). Based on the estimated propensity score, ten subclasses were created with an equal number of students in each subclass (i.e., deciles). The Grade 8 algebra effect was estimated for each subclass based on the same OLS regression models discussed previously. This approach tests for heterogeneity among those who are more or less likely to take algebra, which provides a more refined description of heterogeneity relative to the ATT–ATC dichotomy.

Effect Heterogeneity Across Schools

Given potential variations in Grade 8 mathematics course content and quality across schools, the effect of Grade 8 algebra could vary across schools. A Grade 8 algebra effect averaged across multiple schools may convey little practical meaning if the effect differs dramatically from school to school. As such, it is important to examine whether the estimated effects differ not only across probabilities of student selection, but also across the schools. To estimate school-level variation in the estimated effects, I ran a series of regression models similar to those discussed previously, but allowed the intercept (α_j) and algebra effect slope (δ_j) to vary across schools (j). The hierarchical linear model (HLM; Raudenbush & Bryk, 2002) takes the following form:

$$\begin{aligned} \hat{Y}_{ij} &= \alpha_j + \delta_j Z_{ij} + \beta_1 \mathbf{X}_{1ij} \\ \alpha_j &= \gamma_\alpha + \beta_\alpha \mathbf{X}_{2j} + u_{\alpha j} \\ \delta_j &= \gamma_\delta + \beta_\delta \mathbf{X}_{2j} + u_{\delta j}, \quad \begin{pmatrix} u_{\alpha} \\ u_{\delta} \end{pmatrix} \sim \text{MVN} \left(\begin{pmatrix} 0 & \tau_\alpha \\ 0 & \tau_{\alpha\delta} & \tau_\delta \end{pmatrix} \right), \end{aligned} \quad (10)$$

where the conditional mean outcome value (α_j) and the average algebra effect (δ_j) for school j are factors of an overall grand mean (γ_α and γ_δ , respectively), school-level characteristics (\mathbf{X}_{2j}), and school residuals ($u_{\alpha j}$ and $u_{\delta j}$, respectively). School-level heterogeneity in the average treatment effect is captured by the estimate of τ_α , and the covariance between the conditional school-mean outcome and the conditional school-algebra effect is captured by the estimate of $\tau_{\alpha\delta}$. In the HLM analysis, level-1 covariates (\mathbf{X}_{1ij}) are centered at the respective school-level group means and the level-2 covariates are centered at the grand mean.

Data

Data source. To carry out the analysis I used the NELS. The NELS surveyed over 24,000 Grade 8 students in 1988 through four follow-up data collection periods: 1990 (Grade 10), 1992 (Grade 12), 1994, and 2000. In addition to data on student characteristics, attitudes, and educational activities, the NELS includes data on parents, teachers, and schools. Students were tested in mathematics in the base year and in the follow-up periods. While over 2 decades old, the 1988 NELS cohort is one of the most current nationally representative longitudinal data sources that followed eighth-grade students through high school. The 2002 Educational Longitudinal Study (Ingels et al., 2007) provides more recent nationally representative student data, but starts with a Grade 10 cohort and therefore provides inadequate data for studying Grade 8 selection processes. Conversely, the Early Childhood Longitudinal Study–Kindergarten cohort (ECLS-K; Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009) follows children from their 1998–1999 kindergarten year until their Grade 8 year in 2006–2007, but does not include subsequent

data that would allow one to study the effects of Grade 8 course taking on high school outcomes. While over 20 years old, the proportion of students who took Grade 8 algebra in the NELS analytic sample used for this study is similar to the proportion of students in ECLS-K (Walston & McCarroll, 2010). I elaborate on the similarities of eighth-grade students in NELS and ECLS-K in the Discussion section.

The analysis presented in this article focuses on the NELS base year through second follow-up panel cohort, which followed a subset of the Grade 8 base year sample through Grade 12 ($n = 16,489$). For the analysis, the panel cohort was restricted to students with valid nonmissing information on Grade 8 algebra enrollment and a valid nonmissing Grade 12 mathematics achievement test score ($n = 11,280$). The sample was further restricted to students within a school where at least one Grade 8 student took algebra and at least one student did not take algebra (i.e., the probability of taking Grade 8 algebra was not solely determined by a student's school). The resulting analytic sample included 10,772 students in 941 schools. I also conducted preliminary analyses with other cohort definitions and intermediate outcomes (i.e., Grade 8 and Grade 10 mathematics achievement) but found the substantive findings and conclusions to be the same as those based on the Grade 12 cohort and outcome. Therefore, in this article I only discuss and report the findings that apply to the Grade 12 panel cohort and Grade 12 mathematics achievement.

Weighting the sample. Student data for the NELS were collected using a two-staged stratified probability design, with schools as the primary sampling unit. As a result, analyses that do not take this sampling design into account are likely based on biased standard errors. To account for the NELS sampling design, results presented in this article are based on sampling design weights using the SAS survey estimation procedure whenever appropriate. Note that throughout this article the reported number of observations reflects the unweighted observations.

As discussed previously, inverse propensity weights are also used in the analysis. Results based only on the survey design weights are referenced as design weighted results. Results for the propensity-based weights are also based on the survey estimation procedure. For the ATT estimates, the student-level weights are the ATT weight (see Equation 7) multiplied by the NELS student-level weight (F2PNLWT). For the ATC estimates, the student-level weights are the ATC weight (see Equation 8) multiplied by the NELS student-level weight. Throughout the article, weighted results for the ATT estimates are referenced as ATT weighted and weighted results for the ATC estimates are referenced as ATC weighted.

Treatment variable. The treatment measure of interest is whether a student took a formal algebra class or not in Grade 8. The NELS base year student survey asks participants whether they attended, or did not attend, an algebra (or other advanced mathematics) class at least once a week

during the current school year (BYS67C). Students who responded in the affirmative to this question were coded as having attended Grade 8 algebra ($Z = 1$) and all other students were coded as not attending Grade 8 algebra ($Z = 0$). Overall, about 40% of the panel sample students attended Grade 8 algebra. This dichotomized measure of exposure to Grade 8 algebra was used to distinguish between treatment and control students, and is the dependent variable in the logistic regression model used to construct the propensity scores.

Dependent variable. The outcome of interest for this article is student mathematics achievement at the end of Grade 12. To measure mathematics achievement I used the standardized scores for the NELS second follow-up mathematics test (F22XMSTD). This measure, originally scaled to have a mean of 50 and standard deviation 10, captures mathematics achievement through 4 years of high school. While some studies look at Grade 10 achievement, achievement growth over time, or even number of mathematics courses taken in high school, for simplicity of presentation this study only examined the effect of Grade 8 algebra on end-of-high school mathematics achievement as measured by the NELS mathematics test. For the panel sample, scores on this test ranged from 30 to 71 with a mean of 52.0 and standard deviation of 9.8.

Covariates. The analysis used a number of covariates to account for the potentially confounding factors discussed previously. These covariates include three student characteristics, six family characteristics, seven student academic history indicators, and five school-level characteristics. Summary statistics for the covariates, as well as the dependent and independent variables, are included in the Appendix (Table A1). Taken together, the student and family characteristic measures provide indicators for the nonacademic factors (e.g., cultural and political) theorized to influence selection. The student characteristics include gender, race/ethnicity, and whether a student was limited English proficient. All these measures were dichotomized for the analysis. The family characteristics include: number of siblings, highest parental education level, parental school involvement, a NELS composite of family SES, and indicators for single-parent families and English-only families. The measure of parental education was constructed to minimize missing data by taking the highest level of educational attainment among the parents instead of including separate measures for each parent. A proxy for parental school involvement is utilized based on the following NELS base year survey questions: How often, since the beginning of the school year, have the student and parent(s) discussed selecting courses at the school. The response options for this question were not at all, once or twice, or three or more times, but I dichotomized the measure to distinguish between any discussion and no discussion. The NELS SES measure is a composite of family characteristics including family income, parental occupation, education, and family composition.

Given the academically selective nature of enrollment in Grade 8 algebra, it is important to capture each student's academic history. Unfortunately, the NELS does not include test-based measures of mathematics proficiency prior to Grade 8. One is, however, able to construct a number of indicators for overall academic prowess. While these measures were collected well into the Grade 8 year, they should primarily reflect academic achievement and engagement prior to selection into Grade 8 algebra. The academic history measures included whether the student participated in gifted or talented education (GATE) services, whether the student was ever retained in a grade level, self-reported mathematics class grades during middle school (coded mostly As, mostly Bs, and so on), whether the student planned to attend college, whether the student was ever sent to the office or had parents notified for misbehavior, whether the student was ever sent to the office because of problems with school work, and whether the student's parents ever received a warning about attendance problems. All these measures were dichotomized for the analysis.

To capture school-level factors that may influence selection and outcomes, some school-level characteristics were included in the analysis. These characteristics included whether the student attended a private or public school, urbanicity, school size, percentage of students eligible for the free or reduced-price lunch program, and the percentage of students taking remedial mathematics. For urbanicity, schools were classified as urban, suburban, or rural. For school size, I constructed three categories based on reported Grade 8 enrollment: small (<100), medium (100–299), and large (>299). Similarly, I constructed three categories for the reported percent of eighth-grade students eligible for free or reduced-price lunch: low (<6%), medium (6%–30%), and high (>30%). While the survey did not include a school-level measure for the percent of eighth-grade students taking algebra it did ask for the percentage of students receiving remedial mathematics services. Based on the distribution, I categorized this field into zero, 1%–10%, 11%–20%, and greater than 20%. All these measures were dichotomized for the analysis.

Approximately 9% of the sample had missing data for at least one of the covariates, but the percent of missing data for any one covariate did not exceed 3% (middle school mathematics grades). For covariates missing less than 1% of the data, I imputed the median value. For covariates missing 1% or more of the data, all of which were categorical variables, I included a missing data indicator flag as a separate category. The missing data indicators were included in the analytic models but are excluded from the tables presented in this article.

Results

The Estimated Propensity to Take Algebra in Grade 8

As discussed previously, one method for drawing causal inferences from observational data is to equate the treatment

and control groups based on the propensity of treatment selection. Under the assumption of strong ignorability, valid causal effect estimates are facilitated if the propensity score distribution for the treatment and control groups overlap and, once conditioning on the propensity score, the treatment and control groups exhibit similar covariate characteristics. An estimated propensity score for each student was derived from a logistic regression model (see Equation 6). Because the purpose of the propensity score is to equate the sample treatment and control groups, and not draw population inferences about selection, sample weights were not employed in the propensity score estimation.

Figure 1 shows the distribution of treatment and control group students by propensity score deciles. The distributions do overlap, but as expected, the distribution for the treatment group is concentrated at the upper deciles and the distribution for the control group is concentrated at the lower deciles. For example, 21% of the students who took Grade 8 algebra had a propensity score in the top decile (an estimated propensity score above 0.80), while only 2% of the nonalgebra eighth-grade students had a propensity score that high. Conversely, 15% of the nonalgebra eighth-grade students had a propensity score in the bottom decile (an estimated propensity score below 0.15), while only 3% of the Grade 8 algebra students had a propensity score that low. The key factors for Grade 8 algebra enrollment are apparent when looking at which students fall into the different propensity score deciles. For example, students in the bottom decile are, on average, C students with a history of misbehavior in school, do not want to go to college, and have a parent with no more than a high school education. At the other extreme, students in the top decile are, on average, A students in GATE, want to go to college, and have a parent with at least a college education. The limited overlap between algebra and nonalgebra students in the low and high propensity score deciles reflects the highly selective nature of Grade 8 algebra and suggests estimates of causal effects must take this selective process into account.

While the interest is in the estimated propensity score and not individual coefficients, it is still informative to examine what factors exhibit a strong relationship with Grade 8 algebra enrollment. The logistic regression results are presented in the Appendix (Table A2). As expected, the likelihood that a student takes algebra in Grade 8 is strongly associated with a student's academic history. Students identified for GATE are more than five times as likely to take Grade 8 algebra as students not in GATE, everything else equal. Similarly, receiving mostly As or Bs in middle school mathematics and wanting to go to college are positive predictors of taking Grade 8 algebra. Conversely, students who were retained in a grade at some point in the past and students with a history of misbehaving are less likely to take algebra in Grade 8.

As predicted by the theories that emphasize cultural, political, and organizational factors, student and family characteristics are also associated with algebra placement independent of a student's academic record. For example, students

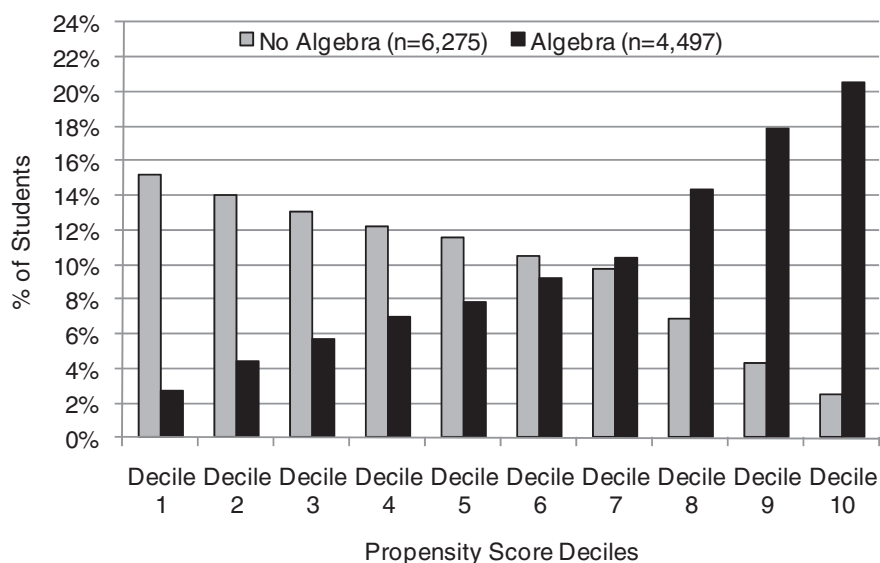


FIGURE 1. Distribution of propensity score for Grade 8 algebra selection across propensity score deciles, by observed selection (algebra or no algebra). Propensity score based on unweighted logistic regression model (see Table A1).

with a college educated parent, high family SES, and parents involved in course selection are more likely to take Grade 8 algebra. Some school factors also seem important. Students in private schools have about a 50% greater likelihood of taking Grade 8 algebra than students in a public school, everything else equal, and students in rural schools are less likely to take Grade 8 algebra.

As discussed in the previous section, to account for observed covariate differences between algebra and nonalgebra students I employed two different propensity score methods. The first method was to reweight the data based on the estimated propensity score to estimate ATT effects and ATC effects (Morgan & Todd, 2008). The second method was to estimate treatment effects within propensity score subclasses (i.e., deciles). The intent of both methods is to create groups that do not differ among the observed covariates (i.e., achieve balance on the observed variables). A common method for assessing covariate balance is to examine the standardized bias reduction after employing the propensity score methods. The standardized bias for each covariate was calculated based on the absolute difference between the treatment and control group means divided by the control group standard deviation (Rosenbaum & Rubin, 1985).

The propensity score methods greatly reduce the amount of bias between the two groups. Group differences and bias among the covariates are reported in the Appendix (Table A3) broken down by the propensity score weighting method. Before weighting the sample based on the propensity score, the average covariate bias between the Grade 8 algebra and nonalgebra groups was 0.17 of a standard deviation. With the ATT-weighted groups the average covari-

ate bias dropped to 0.01 of a standard deviation and with the ATC-weighted groups the bias was 0.02. The most bias reduction occurred among the key covariates. Figure 2 illustrates this bias reduction for four key covariates: college-educated parents, GATE, mostly As in mathematics, and private school attendance. For example, before weighting the sample, the percentage-point difference between treatment and control students with college educated parents was about 22 points. With the ATT weight, the control group is weighted to reflect the treatment group and the percent with college educated parents is virtually identical at about 49%. With the ATC weight, the treatment group is weighted to reflect the control group and the percent with college educated parents is almost identical at about 27%. Similar balancing occurred through subclassification based on the propensity score deciles, where treatment and control group characteristics were similar within each subclass.

Grade 8 Algebra Effect Heterogeneity Across Treatment and Control Groups

With a better understanding of potential selection bias and the balancing role of the propensity score, it is possible to now turn to the primary questions of interest. Does taking algebra in Grade 8 increase a student's mathematics achievement in high school and are some students more likely to benefit from Grade 8 algebra than others? To address these questions I first focused on propensity score weighting methods and the regression-based diagnostics discussed in Morgan and Todd (2008). Students who took algebra in Grade 8 had an average Grade 12 mathematics achievement score of 57,

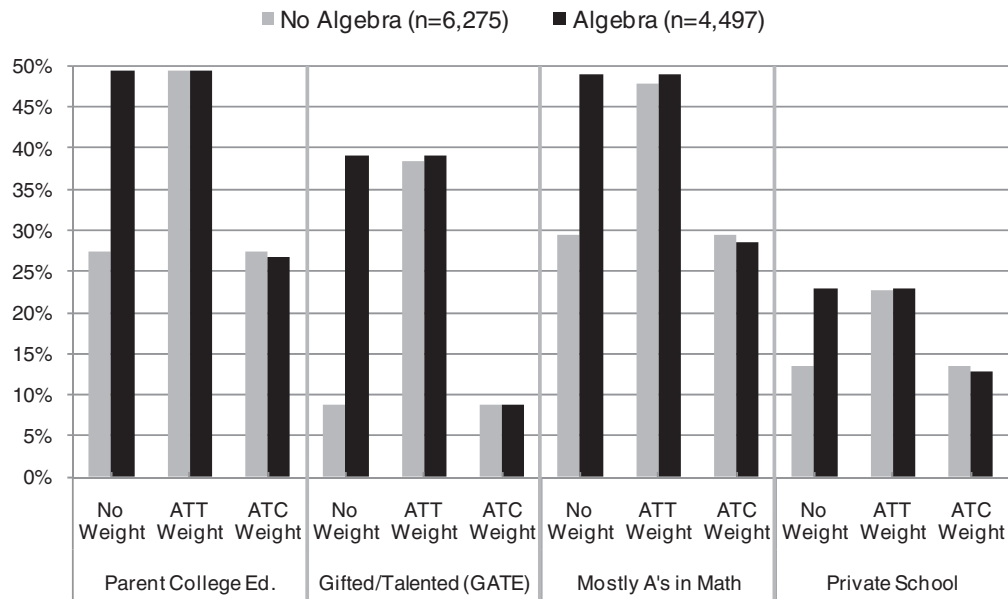


FIGURE 2. Bias reduction for key covariates based on inverse propensity score weighting.

while those who did not take algebra in Grade 8 had an average score of 49. To the naive untrained eye, Grade 8 algebra increases achievement by 8 points, or (with a standard deviation of about 10) by about 0.8 standard deviations. This is a very large effect, but once selection bias is taken into account, this unconditional difference shrinks.

Table 1 presents average effect estimates of Grade 12 mathematics achievement based on different regression models and weighting designs. The design weight column shows how, under a regression-based approach with the survey design weights, the estimated effect shrinks from the unconditionally observed 8.2 estimate (Model 1) to 6.4 with limited covariates (Model 2) to about 3.6 with the full set of covariates (Model 3). The second column of data reports the estimated effects based on the ATE weights, or the expected effect for the average student in the population. Even with the unconditional model (Model 1), the propensity score weighting results in an effect estimate similar to the full conditional model without the propensity score weights. Adding covariates to the model (i.e., moving from Model 1 to Model 3) results in insignificant changes in the effect estimate but, like an analysis of covariance design under random assignment, reduces the standard error. The third column of data reports the estimated effects based on the ATT-weighted approach, or the expected effect for the average student who actually took algebra in Grade 8. The last column of data reports the estimated effects based on the ATC-weighted approach, or the expected effect for the average student who did not take algebra in Grade 8. If effect heterogeneity between treatment and control students exists, then the ATT and ATC estimates should be appreciably different. The results, however, show a relatively steady

treatment effect estimate of about 3.5 across the different propensity score weighting methods and models. Therefore, the effect of Grade 8 algebra appears relatively homogeneous across treatment groups.

Grade 8 Algebra Effect Heterogeneity Across Propensity Score Subclasses

The consistency in effect size across the ATT- and ATC-weighted estimates may still mask some heterogeneity based on the characteristics that make students likely to take Grade 8 algebra. One way to examine this heterogeneity is to estimate the Grade 8 algebra effect within propensity score deciles. Figure 3 displays the estimated treatment effect and 95% confidence interval for each propensity score decile. Each subclass treatment effect was estimated by conducting a separate design-weighted regression, with the full covariate specification, for each subclass. Within each subclass, the Grade 8 algebra effect is positive and statistically different from zero, ranging from about 2 to 5. Additionally, no apparent relationship between effect size and propensity score decile exists. This further supports the notion that the Grade 8 algebra effect is fairly homogeneous.

Grade 8 Algebra Effect Heterogeneity Across Schools

Given variation in instructional quality across schools, the effect of Grade 8 algebra may differ across schools. For example, taking algebra in Grade 8 may not provide as much of a benefit if instructional quality is low relative to Grade 8 algebra classes in other schools. An HLM model (see Equation 10) was employed to test for heterogeneity in the algebra

TABLE 1. Estimated Effects of Grade 8 Algebra on Grade 12 Mathematics Achievement Score, by Weighting Method and Regression Model Complexity

	Design weight		ATE weight		ATT weight		ATC weight	
	Effect	SE	Effect	SE	Effect	SE	Effect	SE
Model 1	8.222**	0.310	3.351**	0.429	3.687**	0.410	3.143**	0.460
Model 2	6.404**	0.279	3.599**	0.314	3.653**	0.320	3.527**	0.344
Model 3	3.584**	0.240	3.568**	0.248	3.603**	0.250	3.547**	0.279

Note. Model 1 does not include covariates; Model 2 only includes covariates for gender, ethnicity, and family socioeconomic status; Model 3 includes full set of covariates discussed in text. See text for discussion of weights. ATE = average treatment effect; ATT = average treatment effect on the treated; ATC = average treatment effect on the control.

** $p < .01$.

effect across schools. I ran three HLM models that allowed the intercept and treatment effect (slope) to vary across schools. All three models included the full set of student-level covariates. The first model excluded school-level covariates (the unconditional model), the second model only included school-level covariates for the intercept, and the third model included school-level covariates for the intercept and treatment effect (conditional model).

Results for the main parameters of interest are presented in Table 2. Across the three HLM models, the overall treatment effect estimate is just over 4. This estimate is slightly higher than the weighted regression estimates presented previously, which is likely a by-product of excluding survey design weights in the HLM models and other differences in the HLM and least squares estimation method. Adding school-level covariates into the HLM model (moving from Model 1 to Model 3) did not appreciably alter the overall

treatment effect estimate. In fact, the only statistically significant school-level covariate for predicting the treatment effect was school size, where small schools were estimated to have about a 1.80 smaller algebra effect relative to medium sized schools.

The overall treatment effect does mask a significant degree of school-level variance in both the intercept (school mean Grade 12 achievement) and the Grade 8 algebra effect. In the unconditional model (Model 1), the intercept variance component was 25.76 ($SD = 5.08$) and the treatment effect variance component was 6.63 ($SD = 2.58$). In the conditional model (Model 3), the intercept variance component was 15.61 ($SD = 3.96$) and the treatment effect variance component is 5.84 ($SD = 2.41$). The school-level covariates accounted for about 12% of the school-level variance in the treatment effect. To get a better sense of the school-level variance in the treatment effect, I calculated

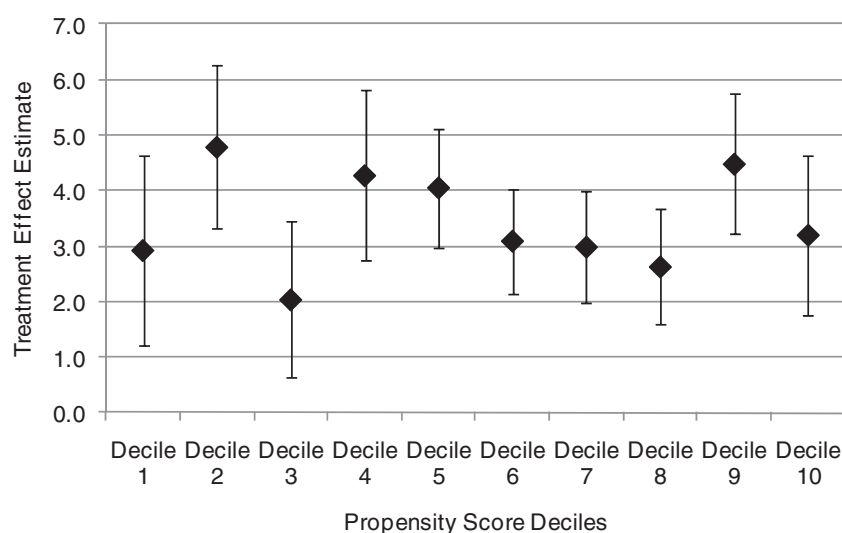


FIGURE 3. Estimated Grade 8 algebra effect on Grade 12 mathematics achievement score across propensity score deciles (subclassification). The error bars represent ± 2 SE.

TABLE 2. Hierarchical Linear Model Estimated Effects of Grade 8 Algebra on Grade 12 Mathematics Achievement Score, by Model Specification

	Model 1		Model 2		Model 3	
	Estimate	SE	Estimate	SE	Estimate	SE
Fixed effects						
Intercept (γ_{α})	51.932**	0.180	51.970	0.147	51.970**	0.147
Treatment effect (γ_{δ})	4.347**	0.188	4.330	0.188	4.242**	0.186
Variance estimates						
Intercept (τ_{α})	25.764**		15.608**		15.610**	
Treatment effect (τ_{δ})	6.631**		6.623**		5.840**	
Covariance ($\tau_{\alpha\delta}$)	-0.468		0.584		0.527	

Note. All three models include full set of level-1 covariates with random intercept and treatment effect. Model 1 does not include level-2 covariates; Model 2 only includes level-2 covariates for the intercept; Model 3 includes level-2 covariates for the intercept and treatment effect.

** $p < .01$.

each school's empirical Bayes estimate for the algebra effect (δ_j) based on the unconditional model and the conditional model. The distributions of the school-level effect estimates are presented in Figure 4. Both distributions ranged from about 0 to 9, with the mass of the distributions falling between about 2 and 6. The unconditional model is more tightly distributed around the grand mean, while the school-level covariates in the conditional model resulted in a slightly more dispersed distribution. In both cases, the school-level variation was apparent, but the vast majority of schools still exhibited positive, and substantively meaningful, Grade 8 algebra effects on Grade 12 mathematics achievement.

In summary, these results are fairly robust across the methods employed in this article so long as a rich set of covariates were accounted for either through a regression-based approach or propensity score approach. It should be noted that both the inverse propensity score method and the subclassification method employed in this article are considered doubly robust methods (Ho et al., 2007; Schafer & Kang, 2008) because they combine regression-based covariate adjustment and propensity score methods. Based on these estimates, taking algebra in Grade 8 had a relatively homogeneous effect on Grade 12 mathematics achievement of about 0.3–0.4 standard deviations. While the HLM models provide evidence for effect heterogeneity across schools, the variance still places average effect sizes for the majority of schools within the rough range of 0.2–0.6 standard deviations. The consistent effect estimates across the treated and control groups, as well as across different propensity levels, suggests that neither a positive selection hypothesis nor a negative selection hypothesis holds in this case. The findings are, in fact, more consistent with the algebra for all hypothesis.

Discussion

Past research on Grade 8 algebra supports policies that encourage more students to take algebra prior to entering high

school. The research, however, is primarily dependent on regression-based effect estimates from observational studies, where the ability to go from statistical associations to policy relevant valid causal inferences is complicated by modeling assumptions and numerous threats to validity. I sought to better understand the robustness of the previous research and the confidence policy analysts and decision makers can place in the findings. In particular, the previous analysis addressed the sensitivity of findings to methods for dealing with selection bias and effect heterogeneity.

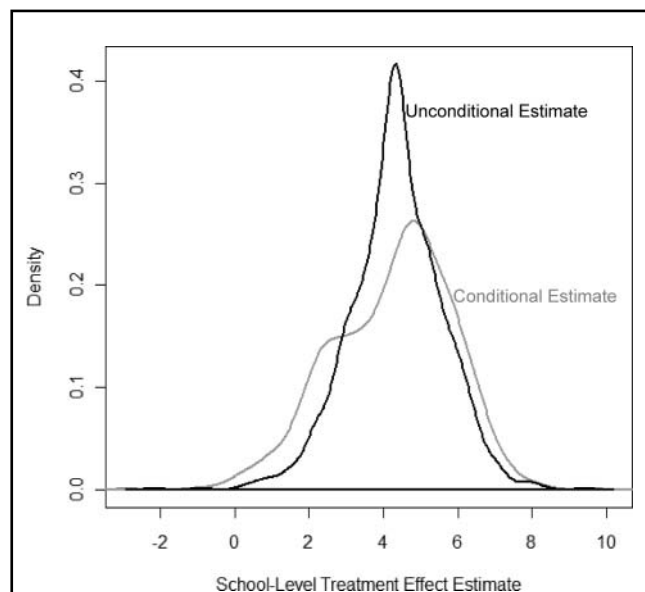


FIGURE 4. School-level kernel density distribution of the hierarchical linear model Grade 8 algebra empirical Bayes effect estimates on Grade 12 mathematics achievement score. The unconditional estimates are based on a model that excluded level-2 covariates. The conditional estimates are based on a model that includes level-2 covariates.

Research on the effects of Grade 8 algebra must be designed to address validity threats due to selection bias. Most students who take algebra in Grade 8 are significantly different from those who take prealgebra. While I found consistent statistically significant positive effects of Grade 8 algebra on long-term mathematics performance, the magnitude of the effect was sensitive to the extent to which the research methods adjusted for potential confounding factors. Pre-existing differences between algebra and nonalgebra students probably accounts for at least half of the observed difference in long-term mathematics achievement, shrinking expected benefits of Grade 8 algebra from about 0.80 standard deviations to about 0.36 standard deviations. Regression-based methods and propensity score methods appear to produce similar results when the same covariates are included in the respective models.

Prior research on the effects of Grade 8 algebra does not directly investigate effect heterogeneity based on selection. Theories related to selection, most notably the positive versus negative selection process theories, imply differential effects between those who take and those who do not take algebra in Grade 8. Additionally, empirical studies suggest some effect differences across ability levels (Gamoran & Hannigan, 2000; Ma, 2005). This article provides a more explicit investigation of effect heterogeneity within the potential outcomes framework. The findings suggest that the effects of Grade 8 algebra are more homogeneous than expected across selection groups. Estimated effects of algebra on Grade 12 mathematics achievement did not differ substantively between students exposed and not exposed to Grade 8 algebra, nor between students with high or low probabilities of taking Grade 8 algebra. Therefore, the results are most consistent with egalitarian efforts to increase access to algebra in middle school, like those expressed by the 2008 National Mathematics Advisory Panel.

While findings from this article support those found in previous research on Grade 8 algebra, the findings are limited by the assumptions imposed in the research design and the data from which they emerge. The potential limitations include breakdowns in the assumptions of strong ignorability and SUTVA, as well as external validity concerns given the restricted sample and data set. I briefly address each of these potential limitations subsequently.

One should not forget that the estimated causal effects assume the results are not affected by unobserved differences between groups after controlling for the observed confounding factors (i.e., the assumption of strong ignorability or conditional independence). This study included a richer set of covariates than some of the previous studies, but was still limited to data collected in NELS. In particular, measures of prior academic performance were limited to student reported grades and a few key gross indicators of academic prowess. Incorporation of detailed information on mathematics achievement prior to Grade 8 would improve confidence in the findings.

I examined how robust the findings are to this assumption in two ways. First, following Frank (2000), I calculated how correlated a key omitted variable—such as a measure of mathematics achievement prior to Grade 8—would have to be with both the treatment indicator and the outcome variable to nullify a statistically significant positive treatment effect. This exercise suggests that the omitted variable would have to have about a 0.60 correlation with both the treatment indicator and outcome variable, independent of the covariates included in the model, to nullify the statistically significant Grade 8 algebra effect on Grade 12 mathematics achievement. None of the observed covariates included in the analysis meet that standard, but it is possible that an omitted confounder does meet that standard. Second, to get a better sense of whether such a strong confounder could exist, I reran part of the analysis (design-weighted Model 3) as if there were a measure of mathematics achievement prior to Grade 8. I had two different assumptions for the properties of this unobserved prior achievement measure. Under one assumption, the prior achievement measure exhibited the same relationships as the observed measure of mathematics achievement at the end of Grade 8. This would be a very conservative estimate given that the observed variable is a posttreatment measure strongly correlated with the treatment and outcome. Under another assumption, the prior achievement measure exhibited the same relationships as the observed measure of science achievement at the end of Grade 8. This would still result in a conservative estimate given the posttreatment measure, but allows for some attenuation because it measures a different content area from the outcome of interest. For both of the assumed prior mathematics achievement measures, the estimated Grade 8 algebra effect remains positive and statistically significant. Under the conservative estimate, the effect size is reduced appreciably to about 0.07 of a standard deviation, but under the less conservative estimate the effect size is still around 0.30 of a standard deviation. This sensitivity analysis suggests the positive Grade 8 algebra effect is relatively robust to omitted variable bias, where only in the extreme case would the effect size be substantively small.

The findings are also based on the SUTVA assumption. This assumption imposes a restriction that assignment of one individual to algebra does not affect the outcomes of another individual, as well as the assumption that the treatment conditions experienced by one individual are the same as those experienced by another individual. Plausible arguments can be made that neither of these assumptions holds under the current policy debate. The idea of one student's course placement not affecting another student's outcomes makes sense when speaking about two average individual students but breaks down when speaking about a policy to shift large numbers of students from one course to the other. Loveless (2008), for example, argued that placing underprepared students in algebra would water down algebra instruction and have a detrimental effect on the learning of well-prepared students. Additionally, the notion that all students

experience roughly the same quality of algebra (or prealgebra) instruction is untenable. Differences in curriculum and teacher quality are likely more important than differences in course titles, thus estimates of potential outcomes for a given individual depends on differences in instructional quality under exposure to algebra or prealgebra. This concern was part of the motivation to examine effect heterogeneity across schools. While I found a significant degree of school-level variation in the algebra effect—presumably due to differences in curriculum, instruction, or student composition of algebra classes—the vast majority of schools still exhibited positive Grade 8 algebra effects.

One argument against expanded access to algebra is the concern that placing certain students in an algebra course before they are prepared to succeed can lead to disengagement from mathematics and ultimately school. Research on algebra effects generally ignores this possible negative effect of dropping out by restricting analyses to students who progress through high school. While the cohort used for this article does contain students enrolled in high school and students no longer enrolled in school (i.e., dropouts), dropouts with nonmissing testing data do not appear to be adequately represented. Thus, understanding the prevalence of any dropout effect has methodological and policy relevance. In particular, the estimated effects may only generalize to students who progress to Grade 12. While not presented, I conducted a preliminary analysis to gauge the algebra effect on dropping out of high school prior to examining the algebra effect on Grade 12 mathematics achievement. I found no evidence that Grade 8 algebra increases the prevalence of dropping out of high school. If anything, the preliminary results suggest that taking Grade 8 algebra might reduce the chance of dropping out. Additionally, preliminary analyses of the Grade 8 algebra effect on Grade 8 and Grade 10 mathematics achievement resulted in findings that parallel those found for Grade 12 achievement. As a result, findings based on analyses focused on 12th-grade students are not likely to be inflated due to lower achieving students dropping out of high school prior to Grade 12, and it is possible to place more confidence in the prior research.

However, one cannot ignore the fact that research on the effects of Grade 8 algebra is primarily based on data collected over two decades ago. While these data sources provide good longitudinal information for studying the long-term implications of educational practices and policies starting in middle school, the assumption that the implications have not changed over the past 20+ years is questionable. It is not clear, for example, how the push for Grade 8 algebra over the past decade (such as California's 1998 adoption of algebra for the Grade 8 content standards) has shifted the distribution of students, resources and quality teachers from prealgebra to algebra classrooms, or, more generally, how these distributions have shifted over time due to demographic, economic, and political changes. Our ability to generalize research findings based on students from the early 1990s to expected outcomes of students in the 2010s de-

pends on the confidence one has in the temporal stability of the educational system. The NELS may be more applicable to the current educational system than one might expect, however, given that the proportion of eighth-grade students taking algebra in the NELS sample and in the ECLS-K 2007 Grade 8 sample are relatively similar. In NELS and ECLS-K, almost 40% of eighth-grade students took algebra and the enrollment rates were similar across key subgroups (rates for ECLS-K are taken from Walston & McCarroll, 2010), such as girls (38% in NELS vs. 42% in ECLS-K), White students (40% in NELS vs. 45% in ECLS-K), students with a college-educated parent (54% in NELS vs. 56% in ECLS-K), and students in a single-parent family (34% in NELS vs. 31% in ECLS-K). Relative stability in mathematics placement for the NELS sample and ECLS-K suggests that the educational experiences of eighth-grade students in the study may be more similar than different from the experiences of today's eighth-grade students.

Despite the previous limitations, this article fills some gaps in the research literature to provide more complete information about whether Grade 8 students are better off in an algebra classroom or a prealgebra classroom. The results support previous research findings that students, on average, can benefit from early access to algebra. More importantly, the results provide evidence that the effect of Grade 8 algebra is fairly homogeneous across students and schools. While this research does not speak to the effects of a universal algebra policy, the mounting evidence supports an educational system where algebra is at least the default Grade 8 curriculum for most students. Changes in curriculum expectations and course titles are not likely to result in substantive improvements in educational outcomes, however, without a better understanding of the underlying instructional and learning mechanisms that make algebra classrooms beneficial for Grade 8 students. Future researchers should strive to uncover how factors such as instructional practices, teaching resources/supports, and classroom contextual effects (e.g., peer effects) mediate the effects of assignment to algebra. The evidence of significant effect heterogeneity across schools indicates that these mediating factors exist, but more in-depth research is required to uncover them. Additionally, efforts to understand how students can best be prepared in earlier grades to succeed in algebra and what supports are necessary to guide underprepared students through algebra are required before it is possible to confidently speak to whether or how all students should take algebra in Grade 8.

ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewers and Michael Seltzer for helpful comments and feedback. Part of this research was made possible by a predoctoral advanced quantitative methodology training grant (#R305B080016) awarded to UCLA by the Institute of Education Sciences of the U.S. Department of Education. An earlier version of this article was presented at the American Educational Research Association 2011 Annual Conference. The views expressed in this article are the author's alone and do not reflect the views or policies of the funding agencies or grantees.

REFERENCES

- Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2009). College preparatory curriculum for all: Academic consequences of requiring algebra and English I for ninth graders in Chicago. *Educational Evaluation and Policy Analysis*, 31, 367–391.
- Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, 30, 51–71.
- Brand, J. E., & Xie, Y. (2010). Who benefits most from college?: Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, 75, 273–302.
- Burris, C. C., Heubert, J. P., & Levin, H. M. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal*, 43, 105–136.
- Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research*, 29, 147–194.
- Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis*, 22, 241–254.
- Gitelman, A. I. (2005). Estimating causal effects from multilevel group-allocation data. *Journal of Educational and Behavioral Statistics*, 30, 397–412.
- Hallinan, M. T. (1990). The effects of ability grouping in secondary schools: A response to Slavin's best-evidence synthesis. *Review of Educational Research*, 60, 501–504.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reduced model dependence in parametric causal inference. *Political Analysis*, 15, 199–236.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–970.
- Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, 101, 901–910.
- Ingels, S. J., Pratt, D. J., Wilson, D., Burns, L. J., Currivan, D., Rogers, J. E., & Hubbard-Bednasz, S. (2007). *Education Longitudinal Study of 2002: Base-year to second follow-up data file documentation* (NCES 2008-347). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Kelly, S. (2004). Do increased levels of parental involvement account for social class differences in track placement? *Social Science Research*, 33, 626–659.
- Lee, V. E., Croninger, R. G., & Smith, J. B. (1997). Course-taking, equity, and mathematics learning: Testing the constrained curriculum hypothesis in U. S. secondary schools. *Educational Evaluation and Policy Analysis*, 19, 99–121.
- Loveless, T. (2008). *The misplaced math student: Lost in eighth-grade algebra. The 2008 Brown Center Report on American Education*. Washington, DC: Brown Center on Education Policy, Brookings Institute.
- Ma, X. (2005). Early acceleration of students in mathematics: Does it promote growth and stability of growth in achievement across mathematical areas? *Contemporary Educational Psychology*, 30, 439–460.
- Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research*, 35, 3–60.
- Morgan, S. L., & Todd, J. J. (2008). A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology*, 38, 231–281.
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research* (1st ed.). Cambridge, UK: Cambridge University Press.
- National Mathematics Advisory Panel. (2008). *Foundations for success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.
- Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal*, 32, 3–33.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79, 516–524.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, 39, 33–38.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688–701.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6, 34–58.
- Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association*, 75, 591–593.
- Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, 13, 279–313.
- Schmidt, W. H. (2004). A vision for mathematics. *Educational Leadership*, 61(5), 1–6.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Educational Research Association.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Slavin, R. E. (1990). Achievement effect of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60, 471–499.
- Smith, J. (1996). Does an extra year make any difference? The impact of early access to algebra on long-term gains in mathematics attainment. *Educational Evaluation and Policy Analysis*, 18, 141–153.
- Spade, J. Z., Columba, L., & Vanfossen, B. E. (1997). Tracking in mathematics and science: Courses and course-selection procedures. *Sociology of Education*, 70, 108–127.
- Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research*, 81, 453–492.
- Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 155–176). Thousand Oaks, CA: Sage.
- Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), combined user's manual for the ECLS-K eighth-grade and K–8 full sample data files and electronic codebooks* (NCES 2009–004). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- U.S. Department of Education. (2009). *National Assessment of Educational Progress (NAEP), 2009 mathematics assessment*. Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.
- Useem, E. L. (1992). Getting on the fast track in mathematics: School organizational influences on math track assignment. *American Journal of Education*, 100, 325–353.
- Walston, J., & McCarroll, J. C. (2010). *Eighth-grade algebra: Findings from the eighth grade round of the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K)* (No. NCES 2010-016). *Statistics in brief*. Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.
- Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25, 659–706.
- Wu, H. (2001). How to prepare students for algebra. *American Educator*, 25(2), 10–17.

AUTHOR NOTE

Jordan Rickles is a Postdoctoral Scholar at the University of California, Los Angeles. His research explores the intersection of multilevel modeling, causal inference, and propensity score methods, with particular interest in the evaluation of educational programs and policies.

APPENDIX
Supplemental Tables

TABLE A1. Summary Statistics for Variables Used in the Analysis (n = 10,772)

	M	SD	Min	Max
Dependent variable: Grade 12 mathematics achievement	52.339	9.803	29.880	71.370
Independent variable: Grade 8 algebra	0.417	0.493	0.000	1.000
Student characteristics				
Female	0.515	0.500	0.000	1.000
Ethnicity: White	0.726	0.446	0.000	1.000
Ethnicity: Black	0.085	0.278	0.000	1.000
Ethnicity: Latino	0.112	0.315	0.000	1.000
Ethnicity: Asian/Pacific Islander	0.061	0.239	0.000	1.000
Ethnicity: Native American	0.017	0.129	0.000	1.000
Limited English proficient	0.020	0.141	0.000	1.000
Family characteristics				
Number of siblings	2.185	1.530	0.000	6.000
Highest parent education: Unknown	0.075	0.263	0.000	1.000
Highest parent education: Less than high school	0.084	0.278	0.000	1.000
Highest parent education: High school	0.258	0.437	0.000	1.000
Highest parent education: Junior college	0.214	0.410	0.000	1.000
Highest parent education: College	0.366	0.482	0.000	1.000
Parent school involvement	0.874	0.332	0.000	1.000
Single-parent family	0.174	0.379	0.000	1.000
Family SES composite	0.007	0.783	-2.414	1.907
English-only family	0.794	0.404	0.000	1.000
Academic history				
Gifted—talented	0.215	0.411	0.000	1.000
Ever retained in a grade	0.117	0.322	0.000	1.000
Mathematics grades: Mostly As	0.376	0.484	0.000	1.000
Mathematics grades: Mostly Bs	0.364	0.481	0.000	1.000
Mathematics grades: Mostly Cs	0.179	0.384	0.000	1.000
Mathematics grades: Mostly Ds	0.044	0.206	0.000	1.000
Mathematics grades: Mostly Fs	0.017	0.130	0.000	1.000
Mathematics grades: Ungraded	0.002	0.041	0.000	1.000
Student wants to go to college	0.841	0.366	0.000	1.000
Ever sent to office for misbehaving	0.297	0.457	0.000	1.000
Ever sent to office for poor school work	0.073	0.260	0.000	1.000
Parents ever notified about attendance	0.082	0.275	0.000	1.000
School characteristics				
Private school	0.174	0.379	0.000	1.000
Urbanicity: Urban	0.259	0.438	0.000	1.000
Urbanicity: Suburban	0.410	0.492	0.000	1.000
Urbanicity: Rural	0.331	0.470	0.000	1.000
% School lunch: Low	0.293	0.455	0.000	1.000
% School lunch: Medium	0.428	0.495	0.000	1.000
% School lunch: High	0.263	0.440	0.000	1.000
School size: Small	0.319	0.466	0.000	1.000
School size: Medium	0.437	0.496	0.000	1.000
School size: Large	0.244	0.429	0.000	1.000
% Remedial mathematics: None	0.324	0.468	0.000	1.000
% Remedial mathematics: 1%–10%	0.440	0.496	0.000	1.000
% Remedial mathematics: 11%–20%	0.151	0.358	0.000	1.000
% Remedial mathematics: >20%	0.072	0.258	0.000	1.000

Note. Statistics based on unweighted sample. Missing data indicators are not shown. SES = socioeconomic status.

APPENDIX (CONTINUED)

TABLE A2. Logistic Regression Model Results for Prediction of Grade 8 Algebra Selection

	Coefficient	SE	<i>p</i>	Odds ratio	95% Confidence interval
Student characteristics					
Female	−0.110	0.047	.019	0.896	[0.818, 0.982]
Ethnicity (White omitted):					
Black	−0.110	0.088	.213	0.896	[0.754, 1.065]
Hispanic	−0.254	0.095	.008	0.776	[0.644, 0.935]
Asian/Pacific Islander	0.406	0.108	.000	1.501	[1.215, 1.855]
Native American	0.155	0.176	.379	1.167	[0.827, 1.647]
Limited English proficient	−0.037	0.167	.824	0.963	[0.694, 1.337]
Family characteristics					
Number of siblings	−0.002	0.015	.898	0.998	[0.968, 1.029]
Highest parent education (high school grad omitted):					
Unknown	−0.119	0.096	.218	0.888	[0.735, 1.073]
Less than high school	−0.104	0.102	.312	0.902	[0.738, 1.102]
Junior college	0.169	0.066	.011	1.184	[1.040, 1.349]
College	0.305	0.073	.000	1.356	[1.176, 1.565]
Parent school involvement	0.220	0.075	.003	1.246	[1.076, 1.442]
Single-parent family	−0.083	0.062	.181	0.920	[0.814, 1.039]
Family SES	0.263	0.046	.000	1.300	[1.188, 1.423]
English-only family	−0.132	0.074	.075	0.876	[0.758, 1.013]
Academic history					
Gifted/Talented	1.671	0.058	.000	5.317	[4.747, 5.957]
Ever retained in a grade	−0.508	0.079	.000	0.601	[0.516, 0.702]
Mathematics grades (mostly Cs omitted):					
Mostly As	0.705	0.067	.000	2.023	[1.774, 2.308]
Mostly Bs	0.369	0.067	.000	1.447	[1.270, 1.648]
Mostly Ds	−0.062	0.133	.643	0.940	[0.725, 1.220]
Mostly Fs	−0.214	0.219	.327	0.807	[0.526, 1.239]
Ungraded	0.583	0.527	.269	1.790	[0.637, 5.033]
Want to go to college	0.417	0.073	.000	1.517	[1.315, 1.751]
Misbehaving	−0.239	0.054	.000	0.787	[0.708, 0.875]
Poor school work	−0.059	0.095	.533	0.943	[0.782, 1.135]
Attendance issues	0.089	0.088	.315	1.093	[0.919, 1.299]
School characteristics					
Private school	0.405	0.092	.000	1.500	[1.251, 1.797]
Urbanicity (suburban omitted):					
Urban	−0.008	0.060	.899	0.992	[0.882, 1.117]
Rural	−0.131	0.057	.022	0.877	[0.784, 0.982]
% School lunch (medium omitted):					
Low	0.098	0.062	.111	1.103	[0.978, 1.245]
High	0.227	0.061	.000	1.255	[1.114, 1.414]
School size (medium omitted):					
Small	−0.162	0.067	.016	0.851	[0.746, 0.971]
Large	−0.186	0.060	.002	0.830	[0.739, 0.934]
% Remedial mathematics (none omitted):					
1%–10%	−0.212	0.053	.000	0.809	[0.729, 0.899]
11%–20%	−0.083	0.074	.259	0.920	[0.796, 1.063]
>20%	0.127	0.098	.196	1.135	[0.937, 1.376]

Note. Missing data indicators are not shown, but were included in the model. *p* Values < .001 were set to .000. SES = socioeconomic status.

APPENDIX (CONTINUED)

TABLE A3. Comparison of Covariate Means for No Algebra and Algebra Groups Based on Propensity Score Weighting

	Not weighted			ATT weighted		ATC weighted	
	No algebra	Algebra	Bias	No algebra	Bias	Algebra	Bias
Propensity score (log odds)	−0.842	0.250	1.184	0.220	0.032	−0.892	0.054
Overall mean standardized bias			0.169		0.012		0.019
Student characteristics							
Female	0.521	0.508	0.027	0.503	0.009	0.521	0.000
Ethnicity: White	0.715	0.740	0.056	0.730	0.022	0.701	0.032
Ethnicity: Black	0.095	0.070	0.087	0.072	0.007	0.100	0.016
Ethnicity: Latino	0.129	0.087	0.127	0.095	0.026	0.136	0.021
Ethnicity: Asian/Pacific Islander	0.041	0.089	0.240	0.089	0.000	0.041	0.003
Ethnicity: Native American	0.019	0.014	0.035	0.013	0.006	0.022	0.024
Limited English proficient	0.023	0.017	0.040	0.019	0.015	0.026	0.022
Family characteristics							
Number of siblings	2.274	2.062	0.136	2.061	0.000	2.310	0.023
Highest parent education: Unknown	0.088	0.057	0.110	0.054	0.010	0.086	0.005
Highest parent education: Less than high school	0.110	0.048	0.198	0.050	0.005	0.113	0.008
Highest parent education: High school	0.306	0.191	0.250	0.191	0.001	0.313	0.015
Highest parent education: Junior college	0.218	0.209	0.022	0.209	0.001	0.217	0.003
Highest parent education: College	0.274	0.493	0.491	0.493	0.000	0.268	0.014
Parent school involvement	0.848	0.911	0.174	0.913	0.005	0.838	0.029
Single-parent family	0.194	0.146	0.123	0.148	0.006	0.197	0.007
Family SES composite	−0.162	0.244	0.545	0.228	0.022	−0.200	0.050
English-only family	0.801	0.785	0.041	0.774	0.026	0.802	0.003
Academic history							
Gifted/Talented	0.089	0.392	1.067	0.384	0.026	0.087	0.004
Ever retained in a grade	0.156	0.064	0.253	0.065	0.003	0.164	0.024
Mathematics grades: Mostly As	0.294	0.490	0.429	0.478	0.027	0.286	0.018
Mathematics grades: Mostly Bs	0.372	0.354	0.037	0.363	0.018	0.370	0.004
Mathematics grades: Mostly Cs	0.224	0.117	0.258	0.119	0.007	0.230	0.015
Mathematics grades: Mostly Ds	0.060	0.022	0.161	0.023	0.002	0.062	0.009
Mathematics grades: Mostly Fs	0.024	0.007	0.113	0.007	0.002	0.019	0.032
Mathematics grades: Ungraded	0.002	0.002	0.005	0.002	0.001	0.002	0.022
Student wants to go to college	0.783	0.922	0.338	0.918	0.010	0.766	0.040
Ever sent to office for misbehaving	0.338	0.238	0.211	0.238	0.001	0.353	0.030
Ever sent to office for poor school work	0.087	0.053	0.120	0.054	0.004	0.100	0.047
Parents ever notified about attendance	0.096	0.062	0.115	0.067	0.015	0.101	0.015
School characteristics							
Private school	0.135	0.229	0.276	0.226	0.008	0.129	0.015
Urbanicity: Urban	0.236	0.293	0.134	0.282	0.025	0.234	0.005
Urbanicity: Suburban	0.398	0.427	0.058	0.423	0.008	0.391	0.014
Urbanicity: Rural	0.366	0.281	0.177	0.295	0.029	0.375	0.018
% School lunch: Low	0.249	0.354	0.242	0.353	0.001	0.240	0.022
% School lunch: Medium	0.455	0.391	0.129	0.390	0.001	0.461	0.011
% School lunch: High	0.282	0.237	0.100	0.237	0.000	0.286	0.009
School size: Small	0.312	0.330	0.039	0.346	0.036	0.333	0.046
School size: Medium	0.439	0.435	0.009	0.420	0.028	0.419	0.041
School size: Large	0.249	0.236	0.031	0.233	0.006	0.248	0.002
% Remedial mathematics: None	0.302	0.354	0.114	0.333	0.048	0.293	0.020
% Remedial mathematics: 1%–10%	0.458	0.415	0.088	0.440	0.051	0.478	0.038
% Remedial mathematics: 11%–20%	0.157	0.144	0.035	0.142	0.005	0.149	0.021
% Remedial mathematics: >20%	0.070	0.074	0.013	0.074	0.001	0.068	0.009

Note. Bias reflects absolute difference between means of the no algebra (control) group and algebra (treatment) group divided by the unweighted control group standard deviation. Average treatment effect on the treated (ATT) weight bias compares weighted control group means to design weighted treatment group means. Average treatment effect on the control (ATC) weight bias compares weighted treatment group means to design weight control group means. See text for discussion of weights. Missing data indicators are not shown. SES = socioeconomic status.

Copyright of Journal of Educational Research is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.