Journal of Research on Educational Effectiveness, 5: 261–289, 2012

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/15348431.2012.688421

Routledge
Taylor & Francis Group

Weighting Methods for Assessing Policy Effects Mediated by Peer Change

Guanglei Hong
The University of Chicago, Chicago, Illinois, USA

Takako Nomi
Consortium on Chicago School Research at The University of Chicago, Chicago, Illinois, USA

Abstract: The conventional approaches to mediation analysis such as path analysis and structural equation modeling typically involve specifying two structural models, one for the mediator and the other for the outcome. We employ an alternative approach that avoids some strong identification assumptions invoked by the conventional approaches. By applying a new weighting procedure to the observed data, we estimate the average potential outcome if the entire population were treated, the average potential outcome if the entire population were untreated, and the average potential outcome if the entire population were treated and if every individual unit's mediator value would counterfactually remain at the same level as it would be when untreated. The estimated differences among these average potential outcomes provide estimates of the total effect, the natural direct effect, and the natural indirect effect. Applying this approach to multilevel educational data, we evaluate the total effect of the algebra-for-all policy in the Chicago Public Schools by comparing the math achievement of two ninth-grade cohorts. We further investigate whether the policy effect was mediated by the policy-induced change in class peer ability. Combining weighting with prognostic score-based difference-in-differences adjustment enables us to reduce both measured and unmeasured confounding.

Keywords: Causal inference, causal mechanism, direct effect, indirect effect, marginal mean weighting through stratification, potential outcomes, propensity score, prognostic score, ratio-of-mediator-probability weighting

This study introduces a new set of analytic procedures for revealing mediation mechanisms in multilevel settings. We apply these procedures in an investigation of a citywide curricular policy change in Chicago. Inference is based on a comparison between one cohort of students who entered the system before the policy was introduced and another cohort after the policy was implemented. We use marginal mean weighting through stratification (MMWS) to adjust for observed between-cohort demographic differences. To adjust for additional observed and unobserved between-cohort differences attributable to concurrent historical changes, we employ a prognostic score-based difference-in-differences approach. The ratio-of-mediator-probability weighting (RMPW) method then decomposes the total policy effect into a direct effect and an indirect effect mediated by class peer ability change. Taking advantage of the multicohort data, we reduce mediator-outcome confounding associated with not only covariates unaffected by the policy but also school-level covariates

Address correspondence to Guanglei Hong, 5736 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: ghong@uchicago.edu

that could have been affected by the policy. A major contribution of the study is the application of the RMPW method to multilevel educational data. This method provides a viable alternative to the standard methods such as path analysis and structural equation modeling (SEM) by simplifying outcome model specification and by allowing the treatment effect to depend on mediator values.

In 1997, the Chicago Public Schools (CPS) introduced a policy that required all students to take algebra by the end of ninth grade. Prior to 1997, whether a ninth grader took algebra primarily depended on the student's math preparation in the elementary school. The algebra-for-all policy was intended to eliminate remedial math courses and thereby improving high school math achievement across the board. As schools increased algebra enrollment, however, they often created mixed-ability algebra classes by enrolling lower ability students in the same classes with higher ability peers. The reorganization of ninth-grade math classes may have unintended consequences for some students. An earlier study has shown that, although the policy indeed increased algebra enrollment among lower ability ninth graders, the policy showed no detectable impact on their math achievement and yet a negative impact on their grades and passing rates (Allensworth, Nomi, Montgomery, & Lee, 2010). In the meantime, for higher ability students who would have taken algebra regardless of the policy, there was a negative impact on their math achievement likely due to a decline in class peer ability (Nomi, 2010).

This evidence seems to indicate that policy-induced changes in class ability composition may have implications for ninth graders' math learning. Class peer ability represents the amount of math knowledge and skills collectively brought by students in a class. In theory, even when the curriculum is given, ability composition of a class may nonetheless influence instructional content, pace, participation structure, peer interactions, and evaluation, which may subsequently influence a student's math learning and relative standing in class. Changes in class ability composition may also lead to a reallocation of instructional resources between and within classes (Harris, 2010). Hence, how much a student would benefit from taking algebra may partly depend on the ability level of classmates. We reason that changes in class composition may mediate the effect of the algebra-for-all policy on students' math outcomes. For lower ability students in particular, experiencing a rise in peer ability may have mixed effects. Being placed in the same class with higher ability peers may heighten peer competition, increase anxiety for failure, trigger frustration and alienation, and lower one's relative standing and self-esteem due to unfavorable social comparisons. Yet participating in academic discourse that involves higher ability peers is expected to advance lower ability students' math learning unless the algebra content is beyond their reach. In light of the earlier evidence, we expect that the mediation mechanism would be different for higher achieving students.

Because the policy changed not only lower ability students' course taking but also their class peer composition, unpacking the overall policy impact on their math achievement is especially challenging. To illustrate, our causal questions focus on decomposing the total policy effect into the indirect effect mediated by class peer composition change and the direct effect of the policy for the subpopulation of lower ability students. Specifically, we ask: (a) Did the increase in class peer ability mediate the policy effect on these students' math achievement? (b) Would the policy have a direct effect on these students' math achievement if their class peer composition had remained unchanged by the policy?

We organize the article as follows: After defining the causal effects and introducing the data, the article provides an overview of the methodological challenges. We show that the conventional approaches require identification assumptions that are apparently implausible in the current application and, we suspect, in many other applications in educational research. We then propose a new approach to mediation analysis, clarify the assumptions under which the causal effects can be identified, and evaluate the assumptions in relation to the data. This is followed by a detailed explanation of the analytic procedure and a report of the empirical results. The last section discusses the strengths and limitations of our proposed methods and raises methodological issues for future research.

DEFINITIONS OF CAUSAL EFFECTS

Total Policy Effect

Let $Z_{ik\cdot} = 1$ if student i in school k attended the ninth grade after the algebra-for-all policy was introduced and 0 otherwise, assuming that the student would attend the same neighborhood school regardless of the policy. The student's math achievement is a function of the policy, denoted by $Y_{ik}(z)$. The student would display potential outcome $Y_{ik}(1)$ if attending the ninth grade after the policy was introduced and would display potential outcome $Y_{ik}(0)$ instead if attending the ninth grade before the policy was introduced. If the student's potential outcome values do not depend on *when* and *how* the treatment was delivered and what treatments were received by other students in the population, under this Stable Unit Treatment Value Assumption (SUTVA; Rubin, 1986), the total effect of the policy for this student is simply the difference between the two potential outcomes: $Y_{ik}(1) - Y_{ik}(0)$. The average total effect of the policy for all students in this subpopulation is defined as $E[Y(1) - Y(0)]$, which is equivalent to $E[Y(1)] - E[Y(0)]$, that is, the difference between the average potential outcome if all ninth graders in this subpopulation would have attended Chicago high schools after the policy was introduced and the average potential outcome if all of them would have been present before the policy was introduced. The former would have been observable had the policy taken effect in an earlier year; whereas the latter would have been observable had the policy been postponed.

Even though a student's school membership is given, when the policy is implemented in classrooms, the student's potential outcome values may depend additionally on the classroom setting (Hong, 2004; Hong & Raudenbush, 2006). As we have theorized earlier, peer ability composition in a math class is an important feature of the classroom setting that may constrain instruction and may affect a student's math achievement. Moreover, class peer ability itself is most likely an immediate result of the policy because many schools created mixed-ability algebra classes in the postpolicy year. Student i in school k might be assigned to algebra class j and experience peer ability denoted by $C_{ijk}(1)$ if attending the ninth grade after the policy was introduced and might be assigned to remedial math class j' and experience peer ability $C_{ij'k}(0)$ if attending the ninth grade before the policy was introduced. In this study, math outcome and class peer ability are both measured on continuous scales. Representing a potential math outcome as a function of policy z and class peer ability $C(z)$, the average total policy effect can be written as $E\{Y[1, C(1)] - Y[0, C(0)]\}$.

Decomposition of the Total Policy Effect

Resorting to the potential outcomes framework, the recent statistics literature on mediation has clarified the conceptual distinctions between controlled direct effects, natural direct effect, and natural indirect effect (Pearl, 2001, Robins & Greenland, 1992). Let $Y[1, C(0)]$ denote a student's counterfactual math outcome under the algebra-for-all policy with class peer ability counterfactually remaining unchanged by the policy. The *natural direct effect* of interest here is the expected change in math outcome associated with the policy that cannot be attributed to policy-induced changes in class peer ability, represented as $E\{Y[1, C(0)] - Y[0, C(0)]\}$. The natural direct effect of the policy would inform us, for example, of how the policy might have affected student outcome if the schools had not created mixed-ability algebra classes. The *natural indirect effect* is the expected change in student math outcome solely attributable to the policy-induced change in class peer ability, represented as $E\{Y[1, C(1)] - Y[1, C(0)]\}$. The sum of the natural direct effect and the natural indirect effect is the total effect. The decomposition of the total effect is not unique because, alternatively, one may define the natural direct effect as $E\{Y[1, C(1)] - Y[0, C(1)]\}$ and the natural indirect effect as $E\{Y[0, C(1)] - Y[0, C(0)]\}$.

In contrast, a *controlled direct effect of the policy*, represented as $E[Y(1,c) - Y(0,c)]$ would be conceivable if another intervention held class peer ability at a fixed level c regardless of whether a student attended the ninth grade before or after the policy was introduced. The controlled direct effect and the natural direct effect would be equal if the controlled direct effect did not depend on the mediator value c. For example, they would be equal if how much a student would benefit from taking algebra did not depend on class peer ability.

This framework of natural direct and indirect effects relaxes SUTVA for potential outcome values $Y[z, C(z)]$ and for potential mediator values $C(z)$. Under the algebra-for-all policy, a student's peer ability and math outcome might conceivably have taken values different from the observed ones if, for example, the school had counterfactually decided to continue the same practice of sorting students to math classes as that prior to the policy, which would lead to a lack of change in class peer ability despite the policy change. In our view, decomposing the total policy effect would have been impossible without relaxing SUTVA.

DATA

Population

The population includes all 59 CPS neighborhood high schools in existence before and after the algebra-for-all policy was introduced in 1997. Among them 14 schools offered algebra to all ninth graders even prior to 1997. We select one prepolicy cohort and one postpolicy cohort of first-time ninth graders who were not receiving special education services.

Math Outcome

Student math achievement score comes from the Tests of Academic Proficiency administered at the end of the ninth grade. The instrumentation did not change across the cohorts.

Class Peer Ability

A student's latent math incoming ability is assessed on the basis of the Iowa Tests of Basic Skills achievement trajectory from the third to the eighth grade. This measure has been standardized across multiple cohorts of students. We use the class median math incoming ability to represent the average peer ability within a class. To ensure reliability of measurement, classes with fewer than five students in the sample or with more than 30% of students missing incoming test scores are excluded from the analysis. Given the finite sample size in a class, the median is preferable to the mean because the former is less sensitive to extreme values.

Covariates

We consider all observed covariates that predict class peer ability when the policy is given, that predict student math outcome when the policy is given, or that predict student math outcome when the policy and class peer ability are both given. These are placed in three categories: student pretreatment characteristics denoted by X, school pretreatment characteristics denoted by $\bar{X}$ as most of these measures are school aggregates of student characteristics, and school posttreatment characteristics denoted by $W(z)$ for $z = 0, 1$. Table 1 shows descriptive statistics for the covariates in each category.

Student pretreatment characteristics. These are student characteristics that could not have been affected by policy exposure. The measures include age, gender, race, socioeconomic status (SES), English language learner (ELL) status in the eighth grade, residence in the attendance zone, residential mobility prior to high school, schooling experience in CPS or elsewhere, and latent math and reading skills upon entering high school. Student socioeconomic measures are constructed by linking the 2000 U.S. census block-level data to student home address.

School pretreatment characteristics. These are school characteristics that could not have been affected by the policy. The measures include the ninth graders' racial composition, age composition, SES composition, freshmen cohort size, percentage of ELL students, percentage of special education students, percentage of students with a history of residential mobility, percentage of students from private elementary schools, percentage of students from non-CPS public schools, percentage of students from the attendance zone, whether the school was vocational, whether the school offered algebra to all ninth graders in the prepolicy year, and school mean and standard deviation of ninth graders' incoming math and reading skills.

School posttreatment characteristics. These are school characteristics, measured in the prepolicy year and again in the postpolicy year, that could change over time as a result of the policy and that could affect how students were assigned to classes and therefore potentially confounding the mediator–outcome relationship. The measures include percentage of regular education ninth graders enrolled in algebra, percentage of special education ninth graders enrolled in algebra, percentage of ninth graders enrolled in higher level math courses, total number of ninth-grade math teachers, number of ninth-grade math teachers new to the school, and within-school variability in ninth-grade math class size.

Unobserved covariates. Concurrent historical changes that could lead to unobserved differences between the pre-policy cohort and the postpolicy cohort are denoted by U_X. Student, class, and school characteristics that could have been affected by the policy yet

Table 1. Descriptive statistics

	Prepolicy[a]		Postpolicy[b]	
	M	*SD*	*M*	*SD*
Student Outcome (Y)				
TAP math scores	22.57	10.91	30.12	11.28
Mediator (M)				
Class median ability (cohort mean adjusted)	−0.61	0.37	−0.25	0.33
Student Pretreatment Covariates (X)				
Incoming math ability	−0.94	0.29	−0.72	0.26
Male	0.46	0.50	0.50	0.50
White	0.08	0.27	0.06	0.23
Hispanic	0.34	0.47	0.38	0.49
Asian	0.01	0.11	0.01	0.10
Old for grade	0.15	0.36	0.23	0.42
Social status index	0.09	0.86	0.06	0.80
Poverty index	−0.07	0.87	−0.15	0.84
From the attendance zone	0.62	0.48	0.67	0.47
Moved once	0.26	0.44	0.28	0.45
Moved twice or more	0.13	0.34	0.13	0.33
Receiving ELL services in 8th grade	0.18	0.38	0.23	0.42
Previously received ELL services	0.19	0.39	0.20	0.40
School Pretreatment Covariates ($\bar{X}$)				
Average incoming math skills	−0.26	0.37	−0.03	0.45
Standard deviations in incoming skills	0.73	0.12	0.85	0.15
% special education students	0.13	0.05	0.18	0.09
% White	0.09	0.12	0.08	0.12
% Asian	0.02	0.04	0.02	0.05
% Hispanics	0.25	0.30	0.27	0.30
% students who are old for grade	0.13	0.03	0.21	0.03
% students who moved once	0.27	0.07	0.27	0.07
% students who moved twice or more	0.12	0.04	0.10	0.04
Average social status index	−0.05	0.56	−0.03	0.55
Average poverty index	0.15	0.64	0.11	0.66
% students from attendance zone	0.51	0.30	0.51	0.32
% students who received ELL services in 8th-grade	0.11	0.14	0.12	0.14
Cohort size in hundreds	4.78	1.76	3.86	1.56
School Post-Treatment Covariates ($W(z)$)				
Algebra enrollment rate (all students)	0.72	0.15	0.99	0.03
Algebra enrollment rate (disabled students)	0.36	0.20	0.92	0.16
Algebra enrollment rate (low-ability students without disability)	0.56	0.19	1.00	0.01
% 9th-grade students enrolled in advanced math	0.01	0.02	0.02	0.03
% new teachers teaching 9th-grade math	0.35	0.18	0.34	0.20
No. of 9th-grade math teachers	12.34	4.96	11.22	4.42
% large classes	0.39	0.23	0.39	0.22
% small classes	0.18	0.16	0.27	0.13

Note. TAP = Tests of Academic Proficiency; ELL = English language learner.
[a]Students $N = 997$; schools $N = 30$. [b]Students $N = 541$; schools $N = 28$.

were not included in our list of observed covariates are denoted by $U_W(z)$. For example, the unobserved covariates may include student motivation and parental involvement.

Subpopulation

As theorized earlier, the policy impact and its mediation mechanism may depend on a ninth grader's prior math ability. We empirically identified, among those attending schools that had offered remedial math in the past, 997 prepolicy students and 541 postpolicy students who would likely experience a rise in class peer ability as well as a change in course taking due to the policy.[1] These students tended to have lower incoming skills than other students in the same cohort. About 42% of these lower ability students enrolled in algebra prepolicy; all of them enrolled in algebra postpolicy. After accounting for the general improvement in ninth graders' incoming skills from the prepolicy to the postpolicy year, we still find a full standard deviation increase in class peer ability on average among these students.

Table 1 compares the prepolicy students and the postpolicy students in this subpopulation with respect to the distributions of their ninth-grade math achievement, peer ability, incoming math skills, and other student-level and school-level characteristics. In general, the average ninth-grade math achievement was much higher in the postpolicy year than in the prepolicy year. A notable change also occurred in student age composition: Postpolicy students were more likely to be old for their grade level than their prepolicy counterparts, likely due to the retention policy instituted in 1996. Among school characteristics that could have been affected by the policy, the proportion of small classes increased notably in the post-policy year.

METHODOLOGICAL CHALLENGES

The theoretical relationships among the algebra-for-all policy, class peer ability, student math learning, and the covariates are represented in Figure 1. Our primary interest is in the policy effect on math learning mediated by policy-induced changes in class peer ability. Valid causal inference is threatened by two major sources of confounding. First, observed pretreatment student characteristics X and school characteristics $\tilde{X}$ along with unobserved pretreatment student and school characteristics U_X may confound the policy effect on math learning, the policy effect on peer ability, and the peer ability effect on math learning. Second, observed posttreatment covariates $W(z)$ and unobserved posttreatment covariates $U_W(z)$ may confound the peer ability effect on math learning. Next we discuss the existing approaches to mediation analysis and the assumptions required.

[1]To empirically identify the lower ability students among those attending policy-affected schools, we specify a model predicting class peer ability in the prepolicy year $C(0)$ as a function of student and school characteristics and apply it to both cohorts. Similarly, a model predicting class peer ability in the postpolicy year $C(1)$ is applied to both cohorts. In addition, we specify a model predicting a student's conditional probability of taking algebra in the prepolicy year $M(0)$ and apply it to both cohorts. A student is identified to be in the lower ability subpopulation if the student would likely experience a change in course-taking ($M(0) < .9$) and a rise in class peer ability ($C(1) - C(0) >$ 0.3 SD).

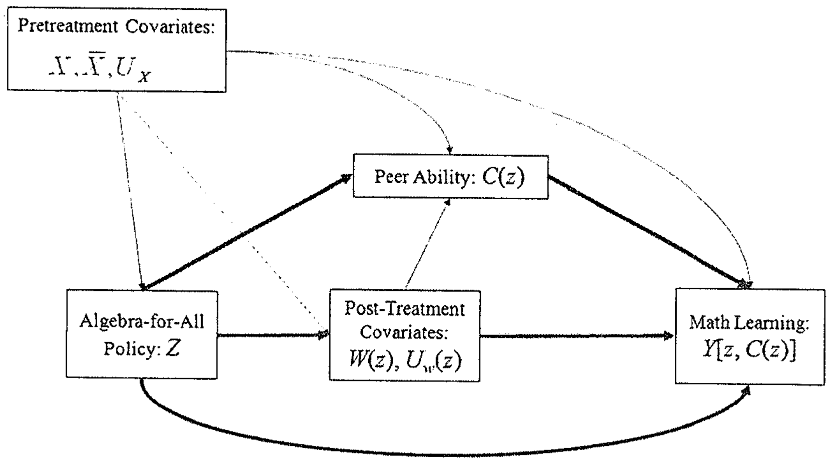

Figure 1. Causal model with potential confounders.

Path Analysis and Structural Equation Modeling

To estimate the natural direct and indirect effects from observed data using path analysis or SEM involves the analysis of two parallel regression models. The first path model regresses class peer ability on policy exposure; the second path model regresses math achievement on class peer ability and policy exposure. Conditioning on the pretreatment covariates, the coefficient for policy exposure in the second model is interpreted as the direct effect of the policy; the product of the coefficient for policy exposure in the first model and that for class peer ability in the second model is interpreted as the indirect effect. The standard error for the estimated indirect effect can be obtained through implementing a Sobel test under distribution assumptions (Sobel, 1982). Identification assumptions required by path analysis and SEM have been explicated in past research (Holland, 1988; Robins & Greenland, 1992; Sobel, 2008) and include the following.

Assumption 1. Nonzero probability of policy exposure conditioning on the observed student and school pretreatment covariates.

$$pr(Z = z \mid \mathrm{X}, \bar{\mathrm{X}}) > 0 \quad \text{for } z = 0, 1.$$

Assumption 2. No confounding of the relationship between policy exposure and math outcome conditioning on the observed student and school pretreatment covariates.

$$Y(z,\ c) \coprod Z \mid \mathrm{X}, \bar{\mathrm{X}} \quad \text{for } z = 0, 1$$

where c takes values from the support for class peer ability within levels defined by X and X̄. An unbiased estimate of the total policy effect on the math outcome can be obtained under Assumptions 1 and 2.

Assumption 3. No confounding of the relationship between policy exposure and class peer ability conditioning on the observed student and school pretreatment covariates.

$$C(z) \coprod Z|\mathrm{X}, \bar{\mathrm{X}} \quad \text{for } z = 0, 1.$$

An unbiased estimate of the policy effect on class peer ability can be obtained under Assumptions 1 and 3.

When Assumptions 1, 2, and 3 hold, policy exposure Z is ignorable conditioning on the observed pretreatment covariates. In the CPS data, even though the algebra-for-all policy was applied to the entire system, the policy effect estimate could possibly be biased by concurrent historical changes. These may include an influx of an immigrant group or the elimination of lowest achieving students in ninth-grade classes by the policy that ended social promotion. Unmeasured historical confounding independent of the observed pretreatment covariates would violate Assumptions 1 to 3.

Assumption 4. Nonzero probability of class peer ability assignment under a given policy conditioning on the observed student and school pretreatment covariates.

$$pr(C(z) = c|Z = z, \mathrm{X}, \bar{\mathrm{X}}) > 0 \quad \text{for } z = 0, 1.$$

It is assumed that, under a given policy z, a student might potentially experience alternative class peer ability levels within the range experienced by those with the same pretreatment backgrounds.

Assumption 5. No confounding of the relationship between class peer ability and math outcome under a given policy conditioning on the observed student and school pretreatment covariates.

$$Y(z, c) \coprod C(z)|Z = z, \mathrm{X}, \bar{\mathrm{X}}.$$

It is assumed that, within levels of the observed pretreatment covariates, class peer ability $C(1)$ is independent of the potential outcome $Y[1, c]$ and that $C(0)$ is independent of the potential outcome $Y[0, c]$.

Under Assumptions 4 and 5, class peer ability assignment under each policy condition is ignorable conditioning on the observed pretreatment covariates. Assumptions 4 and 5 are required for obtaining an unbiased estimate of the effect of class peer ability on the math outcome given the policy. These two assumptions would be violated if there are unobserved pretreatment or post-treatment covariates confounding the mediator–outcome relationship.

Assumption 6. The controlled direct effect of the policy on student math outcome does not depend on class peer ability. In the current application, for those who would take remedial math and therefore experience relatively low peer ability, the controlled direct effect of the policy is likely greater than the effect for those who would take algebra and thus experience relatively high peer ability in the prepolicy year. Hence this assumption would not hold.

Assumption 7. Each treatment effect is constant for all units in the population (Holland, 1988) or, when there is heterogeneity in treatment effects, the individual-specific policy effect on the mediator is independent of the individual-specific mediator effect on the outcome (Bullock, Green, & Ha, 2010).[2] This assumption would be violated, for example, if students who are more susceptible to class peer influence are subjected to a more dramatic shift in class peer ability.

When Assumptions 1 to 7 hold, one can use path analysis, SEM, or similar regression models to decompose the total effect into a natural direct effect and a natural indirect effect by invoking additional model-based assumptions (Baron & Kenny, 1986; Duncan, 1966; Holland, 1988; MacKinnon, 2008). However, if applied to the current study, path analysis and SEM would show limitations due to the implausibility of some key assumptions. Alternative methods have emerged in the statistics literature to relax Assumptions 6 and 7 in particular.

Marginal Structural Models

The regression-based methods as just described cannot handle posttreatment covariates that are potential confounders of the mediator–outcome relationship. For example, percentage of ninth graders taking algebra in a school is a direct result of the policy and is a potential cause of both class peer ability and student math outcome. Adjusting for posttreatment covariates as such through regression would bias the estimation of the total effect and the direct effect of the policy on math learning (Rosenbaum, 1984). This problem can be overcome by applying the marginal structural models (Robins, 2003; Robins & Greenland, 1992; VanderWeele, 2009). Specifically, one may use inverse-probability-of-treatment weighting (IPTW) to approximate a sequential randomized experiment in which students are first randomized to different policy conditions and then, within each policy condition, are randomized to different class peer ability levels. The weight is proportional to the inverse of the conditional probability of class peer ability, the latter being a function of the observed posttreatment as well as the pretreatment covariates. Assumptions 4 and 5 are modified and rendered more plausible. However, to estimate the natural direct effect and the natural indirect effect, this method nonetheless requires Assumption 6.

When the controlled direct effect of the policy depends on class peer ability, neither path analysis/SEM nor marginal structural models apply for estimating the natural direct and indirect effects. Recent research has made important progress in relaxing Assumption 6 by replacing it with other identification assumptions.

Latest Advances in Mediation Analysis

Parametric approaches. A number of alternative methods have been proposed for estimating the natural direct effect or the natural indirect effect within levels of observed pretreatment covariates while allowing the treatment effect to depend on the mediator value and therefore relaxing Assumption 6. They typically involve specifying an outcome

[2]Let a_i denote the policy effect on student i's class peer ability; and let b_i denote the class peer ability effect on math outcome conditioning on policy exposure. Furthermore, let $a = E(a_i)$ be the coefficient for policy exposure in the first path model; and let $b = E(b_i)$ be the coefficient for class peer ability in the second path model. As Bullock et al. (2010) showed, if individual-specific a_i and b_i are not independent of each other, the indirect effect $E(a_i b_i)$ will not be equal to the product of coefficients ab. Instead, we have that

$$E(a_i b_i) = E(a_i) \times E(b_i) + cov(a_i, b_i) = ab + cov(a_i, b_i)$$

model as a function of the treatment, the mediator, the covariates, and their interactions (Pearl, 2010; Petersen, Sinisi, & van der Lann, 2006; VanderWeele & Vansteelandt, 2009). These methods require model-based assumptions with regard to the association between the outcome and the mediator and that between the outcome and the covariates. The functional form of the outcome model may have direct consequences for identification (Drake, 1993; Holland, 1988). Moreover, standard error computation becomes cumbersome for each causal effect estimate represented as a function of multiple parameters.

Less parametric approaches. Other attempts have been made to relax Assumption 6 by taking less parametric approaches. Viewing the counterfactual outcomes as missing data, van der Lann and Petersen (2008) outlined a series of methods for estimating the natural direct effect. The direct effect models require a user-supplied conditional distribution of the mediator representing the data generating function under the control condition. Additional modeling assumptions are necessary for obtaining estimators with good practical performance. Alternatively, Imai and colleagues (Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010) have developed a computationally intensive algorithm that requires fitting a mediator model and an outcome model followed by repeatedly simulating the potential values of the mediator and the potential outcomes given the simulated values of the mediator. The analysis nonetheless depends on correct specifications of both the outcome and the mediator models.

Trade-offs between alternative assumptions. However, unlike the marginal structural models approach, although allowing the treatment and the mediator to interact, the parametric and less parametric innovations require the assumption that, conditioning on the pretreatment covariates, there is no cross-treatment confounding of the mediator–outcome relationship. Assumption 6 is therefore replaced by Assumption 8 (Pearl, 2001; Robins, 2003). Moreover, Assumptions 4, 5, and 8 imply Assumption 7.

Assumption 8. No confounding of the relationship between class peer ability and math outcome across different policies conditioning on the observed student and school pretreatment covariates.

$$Y(z, c) \coprod C(z') | Z = z, \mathrm{X}, \bar{\mathrm{X}}.$$

For example, it is assumed that prepolicy peer ability assignment $C(0)$ is independent of the postpolicy math outcome $Y[1, c]$. When Assumption 8 holds, among those who have the same pretreatment characteristics, the average counterfactual outcome $E\{Y[1, C(0)]\}$ of those who would have experienced peer ability level $C(0) = c$ in the prepolicy year is the same as the average potential outcome $E\{Y[1, C(1)]\}$ of those who actually experienced peer ability level $C(1) = c$ in the postpolicy year. This assumption is violated, for example, if the students who would take remedial math and therefore would experience relatively low peer ability in the absence of the policy would also display a relatively low math outcome when the policy is in place even after controlling for the observed covariates.[3]

In addition, the previous parametric and less parametric innovations require a stronger version of Assumption 2.

[3]Petersen et al. (2006) proposed a mean independence assumption implied by Assumption 8: $E[Y(1, c) - Y(0, c)|\mathrm{X}, \bar{\mathrm{X}}] = E[Y(1, c) - Y(0, c)|C(0) = c, \mathrm{X}, \bar{\mathrm{X}}]$. In other words, the direct effect of the policy no longer depends on class peer ability among those with the same pretreatment characteristics.

*Assumption 2**. No confounding of the relationship between policy exposure and the potential outcomes conditioning on the observed pretreatment covariates.

$$Y(0, C(0)), Y(1, C(1)), Y(1, C(0)) \coprod Z|\mathrm{X}, \bar{\mathrm{X}}.$$

NEW SOLUTIONS

Analytic Strategies

This study introduces a new alternative to mediation analysis with an application to multilevel data. The primary goal is to obtain consistent estimates of the three potential outcomes: $E\{Y[1, C(1)]\}$, $E\{Y[0, C(0)]\}$, and $E\{Y[1, C(0)]\}$, which makes possible the estimation of the natural direct effect and the natural indirect effect. We combine several innovative strategies outlined next.

RMPW estimation of the counterfactual outcome. If students were assigned at random to either the algebra-for-all policy or the control condition, we could easily obtain unbiased estimates of $E\{Y[1, C(1)]\}$ and $E\{Y[0, C(0)]\}$ from the observed data. However, a major challenge to mediation analysis is to estimate the average counterfactual outcome $E\{Y[1, C(0)]\}$. The inference would be possible if the experimental students would counterfactually experience the class peer ability associated with the control condition. Let us suppose that, in an ideal world, students were assigned at random to policy exposure and were subsequently assigned at random to alternative class peer ability levels under each policy condition. We might transform the distribution of class peer ability in the experimental group to resemble its distribution in the control group by assigning the weight

$$\omega_{\frac{c0}{c1}} = \frac{pr(C(0) = c|Z = 0)}{pr(C(1) = c|Z = 1)}$$

to the experimental students who actually experienced class peer ability level c. The weight is a ratio of an experimental student's probability of experiencing peer ability c under the control condition to that under the experimental condition.[4] The weighted mean outcome of the experimental students estimates $E\{Y[1, C(0)]\}$ in this simplified hypothetical experiment. If the assignment to policy exposure were randomized while the assignment to class peer ability were not, we might identify subgroups of students who were homogeneous in the observed covariates. Applying RMPW within each subgroup, the weighted mean outcome of the experimental students would consistently estimate $E\{Y[1, C(0)]\}$ (Hong, 2010b). This strategy is suitable for the current application because it does not require Assumption 6.

MMWS adjustment for observed confounding of policy exposure. Moreover, because policy exposure was not randomized, the pre-policy students and the postpolicy students were different in many important ways. To estimate the average potential outcomes $E\{Y[1,$

[4]For example, in a sequential randomized experiment, suppose that the probability of having high-ability peers is .6 under the experimental condition and .2 under the control condition. To estimate $E\{Y[1, C(0)]\}$, the ratio-of-mediator-probability weight for an experimental student who was assigned to have high-ability peers would be $.2/.6 = 1/3$, whereas the weight for an experimental student who was assigned to have low-ability peers would be $.8/.4 = 2$.

$C(1)]\}$ and $E\{Y[0, C(0)]\}$, we utilize MMWS (Hong, 2010a, 2011) to adjust for between-cohort differences in the math outcome associated with most of the observed pretreatment covariates. The weight is computed nonparametrically on the basis of the propensity of policy exposure. We illustrate MMWS as an alternative to other propensity score-based adjustment methods including matching, stratification, and IPTW.

Prognostic score-based difference-in-differences adjustment for additional confounding of policy exposure. Between-cohort difference in the outcome is partly attributable to other observed or unobserved covariates reflecting the impacts of historical events including concurrent CPS policies. The historical confounding may affect all students including those attending schools that offered algebra to all ninth graders in the prepolicy year and therefore were unaffected by the policy. Modifying the conventional difference-in-differences strategy, we remove the historical confounding locally assessed for subgroups of students who are homogeneous in the observed pretreatment covariates.

Identification Assumptions

Our analyses invoke a set of identification assumptions similar to those required by the less parametric innovations mentioned earlier (Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010; van der Lann & Petersen, 2008). The RMPW strategy does not involve explicit modeling of the mediator–outcome relationship. Hence Assumption 6 becomes unnecessary.

Moreover, given the nature of the multilevel time series data, the same set of schools were present in both the prepolicy year and the postpolicy year. Hence we are able to obtain repeated measures of school characteristics that could have been affected by the policy and could have affected class peer ability and math outcomes. We assume that generally a CPS student would attend the same high school regardless of the year of enrollment. In a given school, prepolicy and postpolicy students would have the same values of school characteristics W(0) if all enrolled in the prepolicy year and would have the same values of W(1) if all enrolled in the postpolicy year. Because policy exposure z is fixed in both W(0) and W(1), their adjustment will not introduce bias.[5] We therefore modify Assumptions 4, 5, and 8 by considering school posttreatment covariates W(0) and W(1) as well as pretreatment covariates X and $\bar{\text{X}}$. This modification increases the plausibility of these assumptions. For $z = 0, 1$ and for c taking values from the support for class peer ability within levels defined by X, $\bar{\text{X}}$, W(0), and W(1), we require that

Assumption 4.* Nonzero probability of class peer ability assignment under a given policy conditioning on the observed student pretreatment covariates and school pretreatment and posttreatment covariates.

$$pr(C(z) = c|Z = z, \text{X}, \bar{\text{X}}, W(1), W(0)) > 0.$$

[5]In general, adjusting for a posttreatment covariate W would introduce bias because $W = W(0)$ is observed only among students in the prepolicy cohort, whereas $W = \text{W}(1)$ is observed only among those in the postpolicy cohort, in which case W is a function of policy exposure Z. In our application, however, $W(0)$ and $W(1)$ are presumably observed for every student. When z takes a fixed value, $W(z)$ is a function of observed and unobserved pre-treatment covariates only and is no longer a function of Z. Hence adjusting for $W(z)$ would not introduce bias in estimating the policy effect.

*Assumption 5**. No confounding of the relationship between peer ability and math outcome under a given policy conditioning on the observed student pretreatment covariates and school pretreatment and posttreatment covariates.

$$Y(z,c) \coprod C(z) | Z = z, \mathrm{X}, \bar{\mathrm{X}}, W(1), W(0).$$

*Assumption 8**. No confounding of the relationship between peer ability and math outcome across different policies conditioning on the observed student pretreatment covariates and school pretreatment and post-treatment covariates.

$$Y(z,c) \coprod C(z') | Z = z, \mathrm{X}, \bar{\mathrm{X}}, W(1), W(0).$$

Next we explain how we employ MMWS and prognostic score-based difference-in-differences to remove historical confounding in attempting to approximate a randomization of policy exposure and, subsequently, how we use the RMPW method to contend with the challenges to mediation analysis. The specific details of the procedure are summarized in Appendix A.

MMWS for Equating Pre-Policy and Post-Policy Student Demographic Composition

As shown in Table 1, the composition of ninth graders changed over time. The MMWS strategy equates the demographic composition of the prepolicy cohort and the postpolicy cohort through propensity score-based weighting. For students attending schools affected by the policy, we summarize the observed demographic information in a propensity score indicating the conditional probability that a student would belong to the postpolicy cohort denoted by

$$\theta_{Z=1} = pr(Z = 1 | \mathrm{X}, \bar{\mathrm{X}}).$$

As described next, potential confounding effects of other historical factors are adjusted through a prognostic score-based difference-in-differences strategy.

We make MMWS adjustment for students in the target subpopulation who attended schools affected by the policy. To satisfy Assumptions 1, 2, and 3, students who did not have counterparts in the alternative cohort are excluded from the analytic sample. The common support is bounded by the logit of propensity scores that are 0.2 standard deviations below

$$max[min(logit_{\theta_{Z=1}} | Z = 1), min(logit_{\theta_{Z=1}} | Z = 0)]$$

and 0.2 standard deviations above the minimum of the two maxima (Austin, 2011). After dividing this analytic sample into $m = 1, \ldots 8$ strata on the basis of the estimated $\theta_{Z=1}$, we compute the marginal mean weight such that the weighted prepolicy cohort and the weighted postpolicy cohort have the same propensity score distribution. The weight for a student with exposure to policy z in stratum m is

$$\omega_z = \frac{n_m \times pr(Z = z)}{n_m^{Z=z}} \tag{1}$$

for $z = 0, 1$. Here n_m is the total number of sampled students in stratum m; and $pr(Z = z)$ is the proportion of sampled students from the cohort that experienced policy z. Hence the numerator represents the number of students in stratum m who would have experienced policy z had policy exposure been completely randomized. The denominator $n_m^{Z=z}$ is the number of students in stratum m who actually experienced policy z.

The two cohorts are expected to display similar distributions of the observed pretreatment covariates after weighting. In general, the prepolicy students and the postpolicy students in this subpopulation showed only small differences in demographic composition. Age (i.e., being old for grade) and ELL status are the only characteristics that differed between the two cohorts by more than 10% of a standard deviation. The weighting adjustment effectively removed these differences as shown in Table 2. Past research has suggested that in the case of a binary treatment, MMWS is as effective as propensity score stratification in removing at least 90% of the selection bias associated with the observed covariates with five or six strata. Yet MMWS is more flexible than propensity score stratification for evaluating multivalued treatments, whereas its nonparametric procedure enhances the robustness of estimation results (Hong, 2010a).

Prognostic Score-Based Difference-in-Differences Adjustment for Other Historical Confounding

We reduce the remaining between-cohort difference in the math outcome associated with other historical confounding by using a non-equivalent comparison group. Fourteen CPS schools offered algebra to all ninth graders even before the policy was introduced. In these schools, class peer ability and student math outcome would not be affected by the policy; hence

$$E\{Y[1, C(1)]\} = E\{Y[0, C(0)]\} = E\{Y[1, C(0)]\}.$$

Any between-cohort difference in average math outcome is to be attributed to historical factors concurrent to the policy. Such information allows us to capture additional historical confounding affecting all schools. Researchers in the past often employ the difference-in-differences method to remove unmeasured historical confounding. This conventional

Table 2. Between-cohort mean differences in student demographic characteristics before and after being weighted by ω_z

Student Demographic Characteristics	Before Weighting	After Weighting
Male	0.04	0.00
White	–0.02	-0.01
Hispanic	0.04	-0.01
Old for grade	0.08	–0.01
Social status	–0.03	0.00
Poverty	–0.08	0.05
Attending school in the attendance zone	0.05	–0.02
Moved once	0.02	–0.01
Receiving ELL services in 8th grade	0.05	0.00
Previously received ELL services	0.01	–0.01

strategy rests on the assumption that the average confounding effect of the historical factors is the same for those unaffected by the policy and those affected by the policy. The assumption would not hold, for example, if the confounding effect is different for Whites and Blacks and if the racial composition differs between the two types of schools. To account for such heterogeneity, we propose to use prognostic score-based adjustment where the prognostic scores are defined as predicted outcomes in the absence of the policy and are functions of the pretreatment covariates. For example, to the extent that race is associated with prepolicy math achievement, the average prognostic scores may differ between Whites and Blacks. The confounding effect estimated as a function of the prognostic scores would then be differentiated between Whites and Blacks.

We intend to identify students in schools affected by the policy and those in schools unaffected by the policy who would display similar prepolicy math outcomes had they attended the same type of schools. Let $G = 0$ if a student attended a school unaffected by the policy and $G = 1$ otherwise. The student's prognostic score $\Psi_{g=0}$ is the predicted prepolicy math outcome if the student would attend an unaffected school; the student's second prognostic score $\Psi_{g=1}$ is the predicted prepolicy math outcome if the student would attend an affected school. The prognostic scores are predicted as functions of the observed pretreatment covariates X and $\bar{\mathrm{X}}$ including student incoming math skills.[6] We allow the pre-treatment covariates to predict the pre-policy math outcome differently across the two types of schools.[7]

$$\Psi_g = E\{Y[0, C(0)]|\mathrm{X}, \bar{\mathrm{X}}, G = g\} = \alpha_1^{(g)}\mathrm{X} + \alpha_2^{(g)}\bar{\mathrm{X}}, \quad \text{for } g = 0, 1. \tag{2}$$

Using data from the unaffected schools, we compute the local confounding effect as a function of the prognostic scores $\Psi = (\Psi_{g=0},\ \Psi_{g=1})$:

$$E[B|\Psi] = E[Y(1)|Z = 1, G = 0, \Psi] - E[Y(0)|Z = 0, G = 0, \Psi] \tag{3}$$

[6]According to Hansen (2008), when the prognostic score is given, the potential outcome associated with the control condition becomes independent of the observed pretreatment covariates. Hence the prognostic score greatly reduces the dimension of covariates to be controlled for. Even though a propensity score estimated as a student's conditional probability of attending an affected school would similarly reduce the dimension of covariates, the prognostic score-based adjustment has important advantages in the current application. Most important, the propensity score for attending an affected school makes no use of the outcome information. In contrast, because the prognostic scores are predicted prepolicy math outcomes and are strongly correlated with the postpolicy math outcomes, conditioning on the prognostic scores, the local confounding effect defined in Equation 3 can be more precisely estimated. Furthermore, the propensity score for attending an affected school would be a function of student characteristics only. The prognostic scores are functions of both student and school characteristics and thus enable us to identify comparable students in comparable schools. Student-level covariates include demographics, incoming skills, social status, poverty status, residence in the attendance zone, mobility, and ELL status prior to high school. School-level covariates include average incoming skills, demographic composition, SES composition (average social status and poverty), mobility rate prior to high school, proportion from the attendance zone, and cohort size in school.

[7]A student's incoming math ability is a stronger predictor of the prepolicy math outcome if attending a school affected by the policy than if attending a school unaffected by the policy. This is because the student would probably take remedial math in a school affected by the policy but would take algebra instead in a school unaffected by the policy.

Table 3. Cell-specific estimate of local confounding effect

Prognostic Score Strata	$E\left[\Psi_{g=0}\right]$	$E\left[\Psi_{g=1}\right]$	$E[B] = E[Y \mid Z = 1, G = 0] - E[Y \mid Z = 0, G = 0]$
1	18.82	18.32	11.76
2	20.09	22.02	8.59
3	21.16	26.00	8.82
4	24.50	22.17	4.46
5	24.55	25.80	8.95
6	25.11	29.26	9.59
7	28.90	25.19	3.66
8	29.35	28.54	8.08
9	29.94	30.92	9.42

Note. Strata 1 through 3 consist of students with $\Psi_{g=0}$ = low, Strata 4 through 6 consist of students with $\Psi_{g=0}$ = medium, Strata 7 through 9 consist of students with $\Psi_{g=0}$ = high. Within each stratum of $\Psi_{g=0}$, students were subdivided into three strata (low, medium, and high) along $\Psi_{g=1}$.

Subtracting the local confounding effect from the postpolicy math outcome of students in the affected schools, we obtain the adjusted local policy effect:

$$\{E[Y(1) \mid Z = 1, G = 1, \Psi] - E[B \mid \Psi]\} - E[Y(0) \mid Z = 0, G = 1, \Psi].$$

This quantity represents the "difference in differences" among those who have the same vector of prognostic score values. The population average policy effect can be obtained by taking integrals over the joint distribution of $\Psi_{g=0}$ and $\Psi_{g=1}$.[8]

We divide the sample into three strata on $\Psi_{g=0}$. Each of these three strata is then subdivided into another three strata on $\Psi_{g=1}$. Table 3 lists the resulting nine cells. The second and third columns show the cell means of $\Psi_{g=0}$ and $\Psi_{g=1}$, respectively. We then obtain an estimate of the local confounding effect for each cell, computed as the cell-specific between-cohort difference in the average math outcome of students in unaffected schools. The estimated local confounding effect, shown in the last column in Table 3, is then subtracted from the postpolicy math outcome of students in affected schools.

In addition to removing unmeasured historical confounding, the prognostic score-based difference-in-differences strategy also enables us to adjust for between-cohort differences in some observed covariates that the MMWS adjustment and other propensity score-based methods are unsuitable for. In particular, system-wide improvement in elementary school education in the 1990s raised math incoming skills of ninth graders by a considerable amount. As shown in Table 1, the average math incoming skills of the lower ability students increased by almost 1 standard deviation. As a result, prepolicy and postpolicy students

[8]We implement this procedure as follows. Analyzing Equation 2, we estimate the coefficients $\alpha_1^{(g=0)}$ and $\alpha_2^{(g=0)}$ using the prepolicy students attending unaffected schools. Similarly, $\alpha_1^{(g=1)}$ and $\alpha_2^{(g=1)}$ are estimated through analyzing the prepolicy students attending affected schools. Next, we apply the model (Equation 2) to predict a pair of prognostic scores for the entire sample. However, the prognostic score model for the pre-policy cohort would not apply to the postpolicy cohort if there were important between-cohort differences in covariate relationships with $Y(0)$. Hence, we apply the marginal mean weight ω_z in all the analysis. Because the between-cohort demographic differences might not be the same across the two types of schools, ω_z is obtained separately for affected schools and unaffected schools.

displaying the same level of math incoming skills differed vastly in their relative standing within the respective cohorts. Moreover, few prepolicy schools could be equated with postpolicy schools on school average math incoming skills. For this reason, we have not included student math incoming skills and the school-level aggregate in the propensity score for policy exposure. Instead, a student's math incoming skills centered at the cohort mean and the cohort mean-adjusted school-level aggregate are among the covariates predicting the prognostic scores. Prepolicy and postpolicy students displaying the same prognostic scores were expected to have the same relative standing in their respective cohorts upon entering high school. The local confounding effect therefore captures the impact of between-cohort difference in the absolute level of math incoming skills.

RMPW for Decomposing the Policy Effect

$E\{Y[1, C(0)]\}$ is the average math outcome that postpolicy students would have displayed had their class peer ability counterfactually remained at the prepolicy level. Under Assumptions 4*, 5*, and 8*, we expect that among those who were homogeneous in the covariates, the counterfactual outcome value $Y[1, C(0) = c]$ would be the same as the observable outcome value $Y[1, C(1) = c]$. Moreover, under Assumptions 1 and 3, the two cohorts would have the same distribution of $C(0)$ within levels of the covariates. Our strategy is to transform the class peer ability distribution in the postpolicy cohort through weighting such that it resembles that in the prepolicy cohort within levels of the covariates. We compute the weight $\omega_{\frac{c0}{c1}}$ for a postpolicy student experiencing class peer ability c as follows:

$$\omega_{\frac{c0}{c1}} = \frac{\phi_{C(0)=c}}{\phi_{C(1)=c}}. \tag{4}$$

After multiplying $\omega_{\frac{c0}{c1}}$ with ω_z, the weighted average math outcome of the postpolicy cohort estimates the counterfactual outcome $E\{Y[1, C(0)]\}$ (Hong, 2010b). See Appendix B for a proof.

To compute $\omega_{\frac{c0}{c1}}$ requires information about a postpolicy student's conditional probability of experiencing class peer ability c that he or she actually experienced during the postpolicy year, denoted by $\phi_{C(1)=c}$, and the student's conditional probability of experiencing class peer ability c had the student attended high school in the prepolicy year instead, denoted by $\phi_{C(0)=c}$. We estimate these two conditional probabilities for each postpolicy student.

Although class peer ability was continuous, we evenly divide the entire population of ninth-grade classes in each cohort, including those attended by higher ability students, into five levels—namely, lowest, lower, medium, higher, and highest denoted by $c = 0, 1, 2, 3, 4$. To adjust for selection of class peer ability under each policy, we estimate two sets of propensity scores—one for the prepolicy year and the other for the postpolicy year—through analyzing an ordinal logistic regression model with students nested within schools in each cohort weighted by ω_z (Zanutto, Lu, & Hornik, 2005). The predictors include all the pretreatment student covariates X, pretreatment school covariates $\bar{X}$, and posttreatment school covariates $W(1)$ and $W(0)$. In addition, the model includes the school-specific random intercept u_0 and the random slopes u_1 for student incoming skills, allowing

Table 4. RMPW transformation of relative frequency distribution of peer ability levels

Peer Ability	Prepolicy (ω_z)		Postpolicy (ω_z)		Postpolicy ($\omega_z \times \omega_{\frac{c0}{c1}}$)	
	Weighted N	Proportion	Weighted N	Proportion	Weighted N	Proportion
Very low	368.23	0.46	68.49	0.15	216.59	0.46
Low	198.80	0.25	147.20	0.32	101.99	0.22
Medium	169.29	0.21	160.20	0.35	96.88	0.21
High	666.68	0.08	87.21	0.19	48.54	0.10
Sum	803	1.00	464	1.00	464	1.00

the selection mechanism to vary across schools.[9] The combined model for each cohort is

$$\eta_c = (\beta_1 + u_1)X + \beta_2\bar{X} + \beta_3 W(0) + \beta_4 W(1) + u_0 + d_c; \quad u \sim N(0, \tau). \tag{5}$$

Here $c = 0, 1, 2, 3$ and $d_0 = 0$; η_c represents the log odds of having peer ability above level c to having peer ability at or below level c. For postpolicy students, the predicted probability $\phi_{C(0)=c}$ can be obtained by applying the coefficients β_1, β_2, β_3, β_4, u_0 and u_1 in Equation 5 that are estimated from the prepolicy cohort data; $\phi_{C(1)=c}$ can be obtained by analyzing the same equation where the coefficients are estimated from the postpolicy data. Because a substantial number of prepolicy students displaying a very high likelihood of experiencing the lowest peer ability did not have counterparts among those experiencing the highest peer ability, we remove the highest peer ability level from the subsequent analysis. Table 4 compares the relative frequency distribution of the four peer ability levels among prepolicy students, among postpolicy students before weighting, and among postpolicy students after weighting.[10] The chi-square test result shows no significant difference in class peer ability distribution between the prepolicy cohort and the RMPW adjusted postpolicy cohort, $\chi^2(3) = 2.15, p > .1$.

To estimate the total policy effect for students in policy-affected schools, a two-level outcome model applies ω_z at the student level. Here the outcome values of students in the postpolicy cohort have been adjusted on the basis of the prognostic scores. In addition, we make covariance adjustment for prognostic score $\Psi_{g=1}$ to improve precision (Yu, 2011). For student i in school j, the outcome model is

$$Y_{ij} = \gamma_0 + Z_{ij}\delta_z + \psi_{g=1,ij}\gamma_1 + u_{0j} + e_{ij}; \quad e_{ij} \sim N(0, \sigma^2); \quad u_{0j} \sim N(0, \tau). \tag{6}$$

[9]To predict prepolicy class peer ability, we used all student covariates and the following school-level covariates: demographic composition, SES composition, percentage of students from attendance zone, percentage of students from non-CPS schools, vocational school, algebra for all students in the prepolicy year, prepolicy algebra enrollment rate, higher math course enrollment rate before and after the policy, algebra enrollment rate among special education students before and after the policy, mean and standard deviation of math incoming skills before and after the policy, and percentage of students in special education. To predict postpolicy class peer ability, we used a similar set of covariates and additionally included the percentage of students receiving ELL services in eighth grade and the percentage of students moved prior to high school.

[10]Because $\omega_{\frac{c0}{c1}}$ is computed directly as a ratio of the two probabilities estimated from the sample, the product of ω_z and $\omega_{\frac{c0}{c1}}$ may become unduly large. We opt for trimming the values larger than 10 by replacing them with a constant 1.0. We then normalize the final weight to restore the sample size.

Under the identification assumptions 1 and 2, δ_z is an unbiased estimate of the total policy effect for students in policy-affected schools.

In the subsequent mediation analysis, we estimate the potential outcome $E\{Y[0, C(0)]\}$ again from the prepolicy data weighted by ω_z. However, to estimate the counterfactual outcome $E\{Y[1, C(0)]\}$ as well as the potential outcome $E\{Y[1, C(1)]\}$ from the postpolicy data requires creating a duplicate set of the postpolicy cohort. After combining this duplicate set with the original sample, we assign the weight as follows:

1. Apply $\omega = \omega_z$ to prepolicy students for estimating $E\{Y[0, C(0)]\}$;
2. Apply $\omega = \omega_z \times \omega_{\frac{c0}{c1}}$ to postpolicy students for estimating $E\{Y[1, C(0)]\}$; and
3. Apply $\omega = \omega_z$ to postpolicy students in the duplicate sample for estimating $E\{Y[1, C(1)]\}$.

Let D be a dummy indicator that takes value 1 for the duplicate postpolicy students and 0 otherwise. We analyze a two-level outcome model weighted by ω:

$$Y_{ij} = \gamma_0 + Z_{ij}\delta_1^{(NDE)} + Z_{ij}D_{ij}\delta_2^{(NIE)} + \Psi_{g=1,ij}\gamma_1 + u_{0j} + e_{ij};$$
$$e_{ij} \sim N(0, \sigma^2); \quad u_{0j} \sim N(0, \tau). \tag{7}$$

Under Assumptions 1 and 2, γ_0 is a consistent estimate of $E\{Y[0, C(0)]\}$, and $\gamma_0 + \delta_1^{(NDE)} + \delta_2^{(NIE)}$ is a consistent estimate of $E\{Y[1, C(1)]\}$. In addition, under Assumptions 3, 4*, 5*, and 8*, $\gamma_0 + \delta_1^{(NDE)}$ is a consistent estimate of $E\{Y[1, C(0)]\}$; therefore, $\delta_1^{(NDE)}$ estimates the average natural direct effect and $\delta_2^{(NIE)}$ estimates the average natural indirect effect. Hypotheses testing are based on robust standard errors that are comparable to sampling variability in Monte Carlo simulation results (Hong, Deutsch, & Hill, 2011).

ANALYTIC RESULTS

Total Policy Effect

Without any adjustment, the mean difference in the math outcome between the prepolicy students and the postpolicy students is 7.82 ($SE = 1.15$, $t = 6.82$, $p < .001$). The analysis is based on data from 1,433 students in 30 policy-affected schools. After applying ω_z that adjusts for between-cohort differences in student and school demographics, the between-cohort difference in the math outcome becomes 7.87 ($SE = 1.29$, $t = 6.10$, $p < .001$). Once we use the prognostic score-based difference-in differences strategy to remove the

Table 5. Estimated total policy effect

Adjustment Type	*Coefficient*	*SE*	*t*
No adjustment	7.82	1.15	6.82***
MMWS	7.87	1.29	6.10***
MMWS and prognostic score-based difference-in-differences	0.25	1.45	0.18
MMWS, prognostic score-based difference-in-differences, and covariance adjustment for prognostic score	0.23	1.15	0.20

***$p < .001$.

Table 6. RMPW estimated natural direct effect and natural indirect effect

	Coefficient	SE	t
Intercept (γ_0)	21.43	0.73	29.54***
Natural direct effect ($\delta_1^{(NDE)}$)	2.70	1.20	2.24*
Natural indirect effect ($\delta_2^{(NIE)}$)	–2.33	0.88	–2.63**

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

confounding associated with additional historical factors including student incoming skills, the between-cohort difference in the outcome is reduced to 0.25 ($SE = 1.45$, $t = 0.18$). Additional covariance adjustment for the prognostic score greatly improves precision. Our final estimate of the total policy effect is 0.23 ($SE = 1.15$, $t = 0.20$). These analyses use data from 1,267 students in 28 schools. Table 5 summarizes the results.

Natural Direct Effect and Natural Indirect Effect

Analyzing the model specified in Equation 7 with the final weight applied at the student level, we decompose the total effect into the natural direct effect and the natural indirect effect mediated by the policy-induced class peer ability change. The estimated natural direct effect is 2.70 ($SE = 1.20$, $t = 2.24$, $p < .05$), and the estimated natural indirect effect is –2.33 ($SE = 0.88$, $t = -2.63$, $p < .01$). These results are summarized in Table 6.

Results Obtained From Path Analysis

For comparison, we apply the conventional procedures outlined in Baron and Kenny (1986). We make covariance adjustment for student math incoming skills and school average income skills and use weighting to adjust for between-cohort difference in student demographics. Moreover, we center the predictors at their respective school means to remove all time-invariant school-level confounding, a strategy equivalent to school fixed-effects analysis (Raudenbush, 2009). The estimated total policy effect on the math outcome is 5.00 ($SE = 1.09$, $t = 4.57$, $p < .001$); the estimated policy effect on class peer ability is 0.57 ($SE = 0.04$, $t = 15.17$, $p < .001$); the estimated direct effect of the policy is 5.39 ($SE = 1.40$, $t = 0.73$, $p = .47$); whereas the estimated effect of peer ability conditioning on policy is –0.94 ($SE = 1.11$, $t = -0.85$, $p = .40$). Hence, the indirect effect of the policy mediated by peer ability change is –0.54 and is not significantly different from zero according to the result of a Sobel test.

CONCLUSION AND DISCUSSIONS

This article introduces a series of weighting in combination with prognostic score-based difference-in-differences in analyzing multilevel cohort data. We evaluated the effect of the algebra-for-all policy on ninth graders' math achievement mediated by policy-induced changes in class peer ability. The subpopulation of interest consists of students who would probably take remedial math instead of algebra in the prepolicy year and were expected to experience an improvement in class peer ability in the postpolicy year. Although the policy did not raise math achievement on average, there is evidence that the policy effect was partly mediated by class peer ability change. The evidence for a negative indirect effect is consistent with the theoretical hypothesis that, for lower ability students, a rise in class peer ability may put them at a disadvantage due to unfavorable social comparisons or because instruction pitched to the middle of the class ability distribution is beyond their grasp. This negative effect may dominate the potential benefit of participating in academic discourse involving higher ability peers. The positive direct effect of the policy, primarily due to the replacement of remedial math with algebra, indicates that exposing lower ability students to algebra would have improved their math learning as intended had their class peer ability remained unchanged.

In general, to reveal the causal mechanism of a system-wide policy is challenging because those who actually experienced the policy might not be identical to those who did not, and because those who experienced different mediator values under a given policy tend to be systematically different. Moreover, there is often a Policy × Mediator interaction effect on the outcome. This study has illustrated how to combine a series of innovative analytic strategies to address possible threats to internal validity. Next we summarize the major strengths of these new solutions and discuss the remaining issues.

Strengths of the New Solutions

RMPW for mediation analysis. As we have explained earlier, the path analysis/SEM approach to mediation (Baron & Kenny, 1986) conventionally assumes that the controlled direct effect of the policy does not depend on mediator values. Recent extensions of these regression methods (Pearl, 2010; Petersen et al., 2006; VanderWeele & Vansteelandt, 2009) relax this assumption typically by invoking model-based assumptions with regard to how the treatment, the mediator, and the covariates interact in the structural model for the outcome. The RMPW method decomposes the total effect into the natural direct effect and the natural indirect effect without imposing model-based assumptions and thus has broader application. According to the results from simulations of single-level data (Hong et al., 2011), under the identification assumptions, the RMPW strategy removes almost all of the initial bias in estimating the natural direct and indirect effects. Generalized least squares analysis generates robust standard errors for both the natural direct effect and the natural indirect effect estimates and therefore provides direct tests of the null hypotheses. Simulation results have shown that the robust standard error estimates are comparable to the simulated sampling variability.

Adjustment for posttreatment covariates. When the direct effect is allowed to depend on mediator values, it is typically required that posttreatment covariates do not confound the mediator–outcome relationship given the observed pretreatment covariates. However, the multilevel cohort data enable us to adjust for posttreatment school covariates. For example, we have repeatedly observed the proportion of small classes in a school. Despite the fact that the policy apparently increased the proportion of small classes, having both prepolicy and postpolicy observations of this covariate for each school allows us to adjust for the selection of class peer ability associated with this posttreatment covariate without introducing bias.

Prognostic score-based adjustment for historical confounding. Because the two cohorts were 3 years apart, the amount of historical confounding could be substantial. Moreover, a major difference between the two cohorts lies in students' math incoming skills due to a concurrent improvement in the CPS elementary schools. By identifying students in policy-affected schools and those in policy-unaffected schools who are relatively homogeneous on a pair of prognostic scores, we are able to locally assess and remove the impact of historical confounding more precisely than does the conventional difference-in-differences approach.[11]

Robustness enhanced by nonparametric weighting. The MMWS method employs a nonparametric procedure in computing the weight that equates the baseline composition of multiple treatment groups. Unlike propensity score-based matching or stratification, the MMWS method is not restricted to evaluations of binary treatments. Past simulation results have shown that in typical applications in which a nonlinear or nonadditive propensity score model is misspecified as a linear and additive one, MMWS estimates of treatment effects display a much higher level of robustness when compared with IPTW estimates (Hong, 2010a).

Precision improved by covariance adjustment for prognostic score. It is well known that weighting increases the variation in estimation. Because the prognostic score summarizes all the observed pretreatment information associated with the outcome under the control condition, adjusting for the prognostic score is equivalent to adjusting for the entire set of covariates with a minimal loss of degrees of freedom. Hence covariance adjustment for the prognostic score in the outcome model effectively improves the precision (Yu, 2011).

Remaining Issues

The strategies that we have proposed here nonetheless require the strong assumptions that policy exposure is independent of potential class peer ability and potential math outcome and that class peer ability under each policy condition is independent of potential math outcome given the observed covariates. Moreover, class peer ability under one policy condition is assumed to be independent of potential math outcome under the alternative policy condition given the covariates. The adjustment for the confounding of mediator–outcome relationship is likely incomplete in the current study. In particular, we were unable to adjust for unobserved pretreatment and posttreatment covariates such as student motivation and teacher quality. Nor were we able to adjust for observed student- and class-level posttreatment covariates such as individual course taking and class size. If the relationship between class peer ability and math achievement could have been confounded by an omitted covariate, sensitivity analysis may be employed to assess the consequence of such an omission (Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010; VanderWeele, 2010).

[11] Covariance adjustment for student incoming skills is seemingly inappropriate because it assumes, for example, that a prepolicy student at the 50th percentile is comparable to a postpolicy student at the 20th percentile displaying the same level of incoming skills. Propensity score-based adjustment shows limitations due to the lack of overlap in the distribution of incoming skills between the two cohorts. As shown in our results, covariance adjustment for incoming skills led to a conclusion that the policy generated a positive total effect on math achievement. In contrast, the result from prognostic score-based adjustment with comparable precision indicates that the average total effect of the policy was indistinguishable from zero.

For example, a student's prepolicy enrollment in algebra is an omitted confounder of the mediator–outcome relationship in the prepolicy year because class peer ability was higher in algebra classes than in remedial math classes and because course taking predicts math achievement. Adjusting for the propensity score for prepolicy class peer ability, we find that course taking no longer predicts math achievement within levels of class peer ability. Hence this omission did not violate the ignorability assumptions. However, we find postpolicy class size to be another omitted covariate that confounds the mediator–outcome relationship in the postpolicy year. Low-ability students in this subpopulation, if attending larger classes, tended to experience higher class peer ability and tended to score lower on the math outcome. Adjusting for the postpolicy propensity score for class peer ability fails to remove this confounding. It could be that, when required to learn algebra, lower ability students were disadvantaged in a large class rather than in a class with relatively high average ability. In the meantime, the policy apparently increased lower ability students' opportunity of attending small classes. Had we been able to isolate a possible positive indirect effect of the policy via class size reduction, we might have detected an even larger negative indirect effect associated with the increase in class peer ability.

The need to examine class size and class peer ability as concurrent and correlated mediators poses a major challenge to causal mediation analysis. Moreover, it is likely that the causal mediation mechanism may vary across schools. Hence another major challenge is to obtain consistent estimates of the variances and the covariance for the random intercept and the random slopes in the final outcome model. We leave these topics for further methodological investigation.

ACKNOWLEDGMENTS

This research was supported by a major research grant entitled "Improving Research on Instruction: Models, Designs, and Analytic Methods" funded by the Spencer Foundation, a Scholars Award from the William T. Grant Foundation and the start-up funds from the University of Chicago for the first author. Additional support came from the Institute of Education Sciences, U.S. Department of Education (Grant No. R305R060059). Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies.

REFERENCES

Allensworth, E., Nomi, T., Montgomery, N., & Lee, V. E. (2010). College preparatory curriculum for all: Academic consequences of requiring Algebra and English I for ninth graders in Chicago. *Education Evaluation and Policy Analysis*, *31*, 367–391.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*, 150–161.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*, 550–558.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*, *49*, 1231–1236.

Duncan, O. D. (1966). "Path analysis: Sociological examples." *American Journal of Sociology*, *72*, 1–16.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, *95*, 481–488.

Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record*, *112*, 1163–1197.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equation models (with discussion). In C. C. Clogg (Ed.), *Sociological methodology* (pp. 449–493). Washington, DC: American Sociological Association.

Hong, G. (2004). *Causal inference for multi-level observational data with application to kindergarten retention* (Unpublished doctoral dissertation). University of Michigan, School of Education, Ann Arbor.

Hong, G. (2010a). Marginal mean weighting through stratification: Adjustment for selection bias in multi-level data. *Journal of Educational and Behavioral Statistics*, *35*, 499–531.

Hong, G. (2010b). Ratio of mediator probability weighting for estimating natural direct and indirect effects. In *JSM Proceedings, Biometrics section* (pp. 2401–2415). Alexandria, VA: American Statistical Association.

Hong, G. (2012). Marginal mean weighting through stratification: A generalized method for evaluating multi-valued and multiple treatments with non-experimental data. *Psychological Methods 17*(1), 44–60.

Hong, G., Deutsch, J., & Hill, H. (2011). Parametric and non-parametric weighting methods for estimating mediation effects: An application to the National Evaluation of Welfare-to-Work Strategies. In *JSM Proceedings, Social Statistics section* (pp. 3215–3229). Alexandria, VA: American Statistical Association.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multi-level observational data. *Journal of the American Statistical Association*, *101*, 901–910.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*, 309–334.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*, 51–71.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Erlbaum.

Nomi, T. (2010, March). *The unintended consequences of an algebra-for-all policy on high-skill students: the effects on instructional organization and students' academic outcomes*. Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Pearl J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* (pp. 411–420). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Pearl, J. (2010, July). *The mediation formula: A guide to the assessment of causal pathways in non-linear models* (Tech. Rep. No. R-363). Los Angeles: University of California, Los Angeles.

Petersen, M. L., Sinisi, S. E., & van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology*, *17*, 276–284.

Raudenbush, S. W. (2009). Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings. *Journal of Education Finance and Policy*, *4*, 468–491.

Robins, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. J. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly structured stochastic systems* (pp. 70–81). New York, NY: Oxford University Press.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*, 143–155.

Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A (General)*, *147*, 656–666.

Rubin, D. B. (1986). Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, *81*, 961–962.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290–312). San Francisco, CA: Jossey-Bass.

Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics*, *33*, 230–251.

van der Lann, M. J., & Petersen, M. L. (2008). Direct effect models. *The International Journal of Biostatistics*, *4*(1), Article 23.

VanderWeele, T. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, *20*, 18–26.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, *21*, 540–551.

VanderWeele, T., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*, 457–468.

Yu, B. (2011). *Variable selection and adjustment in relation to propensity scores and prognostic scores: From single-level to multilevel data* (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Zanutto, E., Lu, B., & Hornik, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics*, *30*(1), 59–73.

APPENDIX A

Analytic Steps

I. Removal of Historical Confounding

Affected by Policy		Unaffected by Policy	
Pre-Policy ($Z = 0, G = 1$)	Post-Policy ($Z = 1, G = 1$)	Pre-Policy ($Z = 0, G = 0$)	Post-Policy ($Z = 1, G = 0$)
Step 1: MMW-S Estimation of $E\{Y[0, C(0)]\}$ and $E\{Y[1, C(1)]\}$			
Estimate propensity score $\theta_{Z=1} = pr(Z = 1 \mid \mathrm{X}, \bar{\mathrm{X}}, G = 1)$ Identify common support on the basis of $\hat{\theta}_{(Z=1\mid G=1)}$ Compute ω_z		Estimate propensity score $\theta_{Z=1} = pr(Z = 1 \mid \mathrm{X}, \bar{\mathrm{X}}, G = 0)$ Identify common support on the basis of $\hat{\theta}_{(Z=1\mid G=0)}$ Compute ω_z	
Step 2: Local Difference-in-Differences Adjustment in the Estimated $E\{Y[1, C(1)]\}$			
Estimate prognostic score under weighting $\Psi_{g=1} = E\{Y(0) \mid \mathrm{X}, \bar{\mathrm{X}}, G = 1\}$ Predict prognostic score $\Psi_{g=0}$	Predict prognostic score $\Psi_{g=1}$ Predict prognostic score $\Psi_{g=0}$	Estimate prognostic score under weighting $\Psi_{g=0} = E\{Y(0) \mid \mathrm{X}, \bar{\mathrm{X}}, G = 0\}$ Predict prognostic score $\Psi_{g=1}$	Predict prognostic score $\Psi_{g=0}$ Predict prognostic score $\Psi_{g=1}$
	Stratify on $\Psi_{g=1}$ and $\Psi_{g=0}$		
		Compute local confounding effects	
	Subtract local confounding effect from post-policy math outcome		

II. Decomposition of the Total Effect

Affected by Policy	
Pre-Policy ($Z = 0, G = 1$)	Post-Policy ($Z = 1, G = 1$)
Step 3: RMPW Estimation of $E\{Y[1, C(0)]\}$	
Estimate pre-policy propensity score model, $\eta_C = \ln\left(\frac{\phi_{C(0)>c}}{\phi_{C(0)\le c}}\right)$	Estimate post-policy propensity score model, $\eta_C = \ln\left(\frac{\phi_{C(1)>c}}{\phi_{C(1)\le c}}\right)$ Estimate $\phi_{C(1)=c}$ Predict $\phi_{C(0)=c}$ Compute $\omega_{\frac{c0}{c1}}$
Step 4: Weighted Estimation of $E\{Y[1, C(0)] - Y[0, C(0)]\}$ and $E\{Y[1, C(1)] - Y[1, C(0)]\}$	
Estimate the total effect weighted by ω_z, with the post-policy outcome already adjusted for local confounding effect	
$\omega = \omega_z$ for pre-policy students	Create a duplicate set of the post-policy cohort Create a dummy indicator for the duplicates; $\omega = \omega_z$ for post-policy students if duplicate; $\omega = \omega_z \times \omega_{\frac{c0}{c1}}$ for post-policy students if original
Estimate the natural direct effect and the natural indirect effect weighted by ω	

APPENDIX B

Let $V = (X, \bar{X}, W(0), W(1))$ represent the observed covariates. We prove that $E(\omega Y|Z = 1)$ consistently estimates $E\{Y[1, C(0)]\}$, where the weight

$$\omega = \frac{pr(C(0) = c|Z = 0, V)}{pr(C(1) = c|Z = 1, V)} \times \frac{pr(Z = 1)}{pr(Z = 1|V)}.$$

The counterfactual outcome is

$$E\{Y[1, C(0)]\} \equiv E\{E(Y[1, C(0)]|V)\}.$$

By Assumptions 1 and 2*, the above is equal to

$$E\{E(Y[1, C(0)]|Z = 1, V)\} \equiv \iiint_{v,c,y} y \times f(Y(z, c) = y|Z = 1, C(0) = c, V = v)$$
$$\times pr(C(0) = c|Z = 1, V = v) \times h(V = v) dydcdv,$$

which, by Assumptions 1, 3, 4*, 5*, and 8*, is equal to

$$\iiint_{v,c,y} y \times f(Y(z, c) = y|Z = 1, C(1) = c, V = v)$$
$$\times pr(C(0) = c|Z = 0, V = v) \times h(V = v) dydcdv$$

which, by Bayes Theorem, is equal to

$$\iiint_{v,c,y} y \times f(Y(z, c) = y|Z = 1, C(1) = c, V = v) \times pr(C(1) = c|Z = 1, V = v)$$
$$\times h(V = v|Z = 1) \times \frac{pr(C(0) = c|Z = 0, V = v)}{pr(C(1) = c|Z = 1, V = v)}$$
$$\times \frac{pr(Z = 1)}{pr(Z = 1|V = v)} dydcdv = E(\omega Y|Z = 1).$$

where

$$\omega = \frac{pr(C(0) = c|Z = 0, V)}{pr(C(1) = c|Z = 1, V)} \times \frac{pr(Z = 1)}{pr(Z = 1|V)}.$$

This concludes the proof.