The Policy Studies Journal, Vol. 36, No. 2, 2008

School Choice as a Latent Variable: Estimating the "Complier Average Causal Effect" of Vouchers in Charlotte

Joshua M. Cowen

Randomized field trials of school voucher policy interventions face major statistical hurdles in the measurement of a voucher effect on student achievement. Selection bias undermines the benefits of randomization when the treatment, a random offer of a voucher, is declined by participants who systematically differ from those who accept. This article argues that the complier average causal effect (CACE) is the parameter of interest in voucher evaluations. As an example, the CACE is estimated using data from a small, one-year field trial of vouchers in Charlotte, NC. In this estimation, voucher impacts in Charlotte are positive, but appear to be moderated by the probability of compliance. For math achievement, maximum likelihood CACE estimates are smaller and insignificant compared to intention to treat and instrumental variable estimates of mean treatment effects.

KEY WORDS: school vouchers, compliance models, causal effects

Few contemporary questions in American education have produced such widespread controversy as that regarding the potential of school vouchers to reduce inequality in student outcomes. This potential has been studied by many scholars of government and public policy, using some of the most advanced scientific research designs whose details are transparent to all critics. These efforts notwithstanding, answers to the voucher question still appear uncertain.

Much of this uncertainty stems from the obstacles faced by evaluators of those few voucher programs, public and private, from which data are available. Of particular difficulty is the accounting of students who chose to enter voucher programs for reasons, perhaps immeasurable, that systematically influence the outcome of interest (typically academic achievement) to program evaluators. The leading scholarship on school vouchers has confronted this difficulty directly, and with some success. In this article, I discuss a particular parameter of interest to school choice evaluators: the complier average causal effect (CACE) of vouchers on student achievement. I argue that because vouchers are inherently a matter of choice, research on the voucher question may benefit from more explicit attention to the different choices made by students when the opportunity to choose is presented, or the lack of choice is imposed. As a concrete example, I examine a small voucher

Published by Blackwell Publishing, Inc., 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford, OX4 2DQ.

program in Charlotte, NC, where test scores were obtained from a sample of voucher users at the end of the 1999–2000 academic year.

1. Background

Some form of a publicly funded voucher program exists today in Milwaukee, WI; Cleveland, OH; Florida; Utah; and the District of Columbia.[1] Privately funded programs have been administered in Milwaukee, the District of Columbia; San Antonio, TX; Indianapolis, IN; New York, NY; Dayton, OH; and Charlotte, NC (Howell & Peterson, 2002). Of these, evaluations of the Milwaukee publicly funded program, and especially the joint evaluation of the privately funded New York, Dayton, and Washington, DC programs have been the most prominent.

Witte (2000) found, at best, mixed results in his evaluation of the impact of the Milwaukee program on student achievement, as gauged by student test scores. Any measurable effects were highly sensitive to the choice of variables, models, sample and statistical tests. Selection and attrition problems undermined the integrity of the randomization achieved by lottery, and even this lottery could not account for differences between applicants to the program and those who did not bother to apply at all. Witte found that other measures of the voucher impact—increased parental satisfaction, for example—were less ambiguous. An unrelated evaluation of the Milwaukee program by Rouse (1998) found that gain scores in math were greater for voucher students than those attending public schools; reading scores, however, changed at similar rates between the two groups. Other work by Greene, Peterson, and Du (1999) was less equivocal in its demonstration of a positive voucher impact in Milwaukee: after accounting for family background, students in voucher schools scored an average 6 to 8 points higher in math, over several years of treatment, and 5 to 9 points higher in reading.

The evaluation of voucher programs in New York, Dayton, and Washington, which culminated in Howell and Peterson (2002), generated considerable admiration for its implementation of a multi-site, multi-year randomized field trial. Some scholars were skeptical of the findings: positive voucher effects on the test scores of privately schooled low-income students. Among the most vocal critics of these results were Krueger and Zhu (2004a, 2004b) who disputed a range of design and analytical choices made by Howell and Peterson (and their colleagues in earlier analyses), most notably the classification of students' race based on the race of their mothers, and the use of baseline test scores as covariates in models of the voucher effect (adding the baseline score limited Howell and Peterson's analysis to students for whom the score was available). However, in a rejoinder to Krueger and Zhu, Peterson and Howell (2004) found little variation in their results even when students without scores were available. In a separate analysis of the New York component of the study, Barnard, Frangakis, Hill, and Rubin (2003) found comparable results to Howell and Peterson. Although the positive results were generally confined to African American students (as opposed to other ethnic groups), and varied across year, city and students' age, the Howell and Peterson studies continue to provide the strongest evidence to date that vouchers can, at minimum, strengthen the central tendency toward improvement among students who use them.

Complicating the results of such research, and that on other education policies, is the issue of selection. Each of the studies noted above is an analysis of data generated by some form of policy trial. Students participating in each study had the option to accept the policy treatment—a school voucher—if offered. Even when the offer was made at random, a nonrandom decision to accept or reject the voucher presented the analysts with a potential selection problem. In general, if nonrandom compliance is related to the same factors influencing the outcome of interest then estimates of the voucher effect are biased.

The literature discussed above contains several attempts to deal with this problem of selection. Largely absent from the literature are explicit discussions of evaluation parameters, including the CACE of voucher treatments. Models of the CACE have been introduced and applied widely in other fields where experimental designs are common, such as biostatistics. In this article, I argue that in addition to addressing a crucial concern in the measurement of a voucher impact on student achievement, the CACE parameter provides great substantive leverage in the description of school choice effects.

2. Compliance in Voucher Experiments

Estimation of the effect of vouchers on student achievement faces a major obstacle. Even in a successful implementation of a randomized field trial for vouchers, researchers must account for nonrandom forces that influence students' decisions to participate and their academic performance simultaneously. A simple estimate of the "choice" parameter—the mean effect of private school attendance on those who use the voucher—is likely biased due to students' nonrandom decision to accept the random voucher treatment. From a policy standpoint alone, it is helpful to determine exactly which factors predict the decline of a voucher offer. Regardless of the net benefit vouchers may hold for those who use them, their promise as a broad-based intervention for students attending failing schools must be at least partially evaluated on the basis of which students might choose to expose themselves to the intervention. From a statistical standpoint, the identification of non-random factors influencing compliance is not merely beneficial but critical to successful evaluation of the policy.

This is not a controversial idea. All of the scholars noted above attempt to model self-selection into private school. Witte (2000), for example, modeled the probability of applying to the Milwaukee Parental Choice Program, conditional on eligibility for the program in the first place, as a function of several demographic characteristics including students' race, family income, gender, and maternal education.[2] He noted that African Americans and Hispanics are more likely to apply to such programs than their white counterparts, girls are more likely to apply than boys, and the probability of applying increases as family income falls and mother's education rises. Prior test scores as predictors of application yielded mixed results. Across their three cities of inquiry, Howell and Peterson (2002) found significantly greater proportions of African American students among students who declined vouchers compared to those who accepted the scholarships. The mothers of decliners also had slightly but statistically significantly fewer years of education. The average income of decliners was, however, statistically significantly higher.[3]

These scholars have applied several techniques centering around instrumental variable (IV) estimation to handle compliance problems in the treatment group. Generally, this approach derives from the work of Heckman (1996) and in randomized field trials employs the voucher offer itself as a variable that, in one stage of the estimation, is correlated with the use of a voucher but (because the offer is random) is uncorrelated with the error in later stage models of student achievement. As such, previous researchers have employed the IV approach to obtain consistent estimates of the average effect of vouchers on the achievement of students who use them (Greene, 2001; Howell & Peterson, 2002). Heckman (1996) notes that this mean provides an answer to the question "how much did persons participating in the program benefit compared to what they would have experienced without participating in the program?" (p. 337). As such, in voucher studies, where private school tuition is a known quantity, this benefit can be weighed against the cost of its implementation and thereby inform policy decisions.

In school choice research, however, this policy decision is complicated not simply by the requirements of addressing self-selection of treatment when it is offered. Added to the analytical burden is the identification of students to whom the policy in question—vouchers, in this case—is likely to appeal as an intervention. As such, even a consistent estimate of the average effect of vouchers on those who use them may not be the most informative outcome of interest.

This is primarily because the implementation of school vouchers is typically a policy targeted to particular segments of the population. Policymakers would benefit from a clear understanding of which factors predict membership in one of two potential groups: one on whom resources may be wasted with vouchers and the other for whom such resources are not enough. In the first group are those students who would always choose a private school, with or without the aid of a voucher. In the second are those who would never make use of a voucher, whether it were offered or not, either because they could not or would not make such a choice. In the interest of effectively targeting a school choice intervention, the relevant group of students from a policy perspective is that whose members, in the absence of a voucher option, stay in public schools but choose private education when the opportunity presents itself. In the context of a voucher field experiment, these students are represented by those participants who comply with whichever experimental condition is assigned to them.

The CACE

Several scholars (e.g., Bloom, 1984; Imbens & Angrist, 1994; Imbens & Rubin, 1997; but for this analysis see specifically Jo & Muthen, 2001; Little & Yau, 1998) have suggested versions of the CACE.[4] This is the mean treatment outcome across the subpopulation of compliers: those subjects who accept the treatment condition assigned. This approach explicitly treats nonrandom noncompliance as a missing data problem, where compliance is a variable whose value is observed for those in the treatment group, but unobserved for the control group. Scholars have applied CACE models widely in such fields as biostatistics (e.g., Dunn et al., 2003; Have et al., 2004; Mealli, Imbens, Ferro, & Annibale, 2004). To my knowledge, however, only one published article (Barnard et al., 2003) explicitly discusses the CACE as a parameter of interest in voucher studies.

CACE models recognize that two subgroups exist among the sample of participants assigned to the control group: one whose members are in practice highly likely (and in theory certain) to seek treatment regardless of experimental condition assigned and the other whose members are sure to reject the offer of treatment. The first group, the "always-takers," (Angrist, Imbens, & Rubin, 1996) attends private schools regardless of condition assignment. Conversely, the second group, the "never-takers" (Angrist et al., 1996) is theoretically identical to the participants who actually are randomly offered the voucher but reject it. If we believe that we need to account for the nonrandom factors that determine a "never-taker" among the sample of participants offered treatment, we should also believe it necessary to account for these factors that would theoretically determine "never-takers" in the control group. Because the CACE is the measure of the treatment impact only for compliers, it is an indication in voucher trials of the impact of school choice on students who "choose" if and only if they actually have the opportunity to do so. In other words, as Heckman (1997) has noted, the CACE can be considered the treatment effect on the treated for subjects at the margins: those whose behavior is explicitly governed by the presence or absence of the intervention itself.

Compliance as a Latent Variable

Bloom (1984) proposed an IV approach to estimating the CACE by weighting the treatment effects by proportions of noncompliers, while others (Imbens & Rubin, 1997; Little & Yau, 1998) have since demonstrated that maximum likelihood estimates are more efficient. This article embraces the flexible framework of structural equation modeling in general, and latent variable mixture models in particular (Little & Yau, 1998; Muthen, 2001), in the application of CACE estimations to the Charlotte voucher data. In this framework, two latent classes,[5] compliers and noncompliers, are used to compare treatment outcomes. Class membership is observed for those offered treatment but unobserved for those in the control group, and maximum likelihood and the EM-algorithm (Little & Rubin, 1987) may be used to estimate parameters of the model predicting class membership. In the case of this study, the model uses the characteristics of those who reject the voucher offer (i.e., the decline group) to predict the latent complier status of those in the control group; a voucher effect is then estimated for all compliers: those whose compliance we observe and those whose compliance is latent.

Figure 1 displays a pictorial representation of a CACE model. Let C represent observed (for treatment students) and unobserved (for control students) compliance, X represent a set of student characteristics, "choice" representing the effect of the

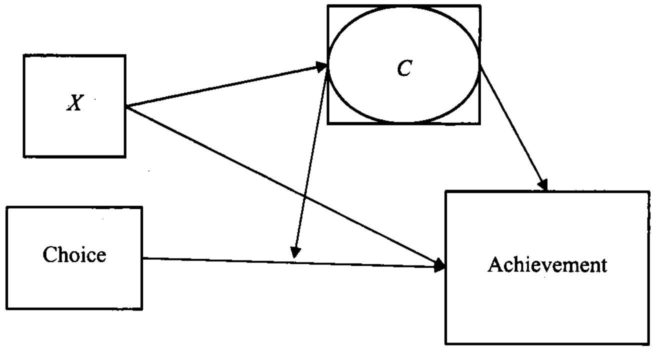

Figure 1. A Complier Average Causal Effect Model of Achievement in Charlotte.

voucher treatment, and "achievement" the outcome (math or reading) of interest. Each arrow represents a relationship between model components. As the figure indicates, the student characteristics, *X*, predict both compliance and (in part) achievement. Mean achievement levels are different for compliers and noncompliers (as the arrow from C to "achievement" indicates), and compliance mediates the treatment effect of the "choice" on achievement itself.

Assumptions. Following the assumptions of CACE estimation detailed in Angrist et al. (1996) and Jo and Muthen (2001) I outline the assumptions of CACE estimation in the latent class framework. The first assumption is randomization of the treatment condition. The second, Stable Unit Treatment Value (SUTVA), is derived from Rubin's causal model (e.g., Rubin, 1974) and holds that the potential outcomes of each study participant are unrelated to the treatment status of other individuals. The third assumption, monotonicity, implies that there are no treatment "defiers" who would do exactly the opposite of whatever treatment condition was assigned (Angrist et al., 1996; Little & Yau, 1998). In the case of school vouchers, in other words, monotonicity assumes that there are no students who would reject the voucher if offered and yet attend a private school if assigned to the control group. The final assumption is known as the exclusion restriction, which implies that the treatment outcome is independent of treatment assignment for the "always-takers" and "never-takers" discussed above.

3. An Example: Estimating the CACE of Vouchers in Charlotte

The Charlotte Scholarship Fund

Data. As I discuss in the first section of this article, field trials of voucher interventions are yet rare. Although rigorous evaluations of the Milwaukee and Cleveland public programs are now contained in the scholarly record, only the evaluation of a private voucher program in Dayton, New York City, and Washington, DC exists in the record as a multi-year, fully implemented randomized field trial. The data from this field experiment are not currently available for public research. As a consequence, a study such as this one, which seeks to place a methodological argument in the literature on voucher effects, must look elsewhere for illustrative data.

To this end, I estimate the CACE of vouchers in Charlotte, NC, where a small-N trial took place during the 1999–2000 academic year. Greene (2000, 2001) details the full design of the voucher trial administered by the Charlotte Children's Scholarship Fund. Here, I include relevant design issues. Incoming second- through eighth-grade students from low-income families in Charlotte were offered the opportunity to apply for a $1,700 scholarship to attend a private school for the 1999–2000 school year. Of the original applicants, 347 (30%) agreed to participate in a program evaluation the following spring. At the end of the school year, Iowa Tests of Basic Skills (ITBS) were administered to all students, while their parents completed surveys designed to obtain background information. There was no pretest. Families who had either lost the lottery or had chosen not to accept the voucher were offered $20 and a chance to win a new scholarship to attend the testing sessions. I report descriptive statistics of the study participants in Table 1.

Nonresponse. Greene (2000, 2001) provided several plausible explanations for the low participation rate. These explanations include outdated contact information, conflicting schedules (the follow-up sessions were held on four different weekends in March and April 2000) and a low financial incentive to participate. Unfortunately, the scholarship providers did not collect baseline information save verification of the low-income eligibility required for all participants in the program as they randomly awarded the money.

This income information, however, is the first indication that whatever the difference between responders and nonresponders, it need not necessarily bias estimates of a voucher effect within the study sample. Although these data limit the generalizability of a voucher effect on achievement in the population at large, the randomized nature of the intervention itself allows estimation of a causal effect

Table 1. Descriptive Statistics

	Choice	Decline	Control	ITT	Total
Math (mean)	36.67	28.54	28.70	34.60	32.30
Reading (mean)	42.71	34.37	33.33	40.58	37.76
African American mother (% = 1)	77.22	92.59**	80.00	81.83	80.69
Two parents (% = 1)	40.51	14.81***	32.59	33.96	33.43
Male student (% = 1)	47.47	44.44	41.48	46.70	44.67
Public assistance (% = 1)	30.38	33.33	25.93	31.13	29.11
Income < $20,000 (% = 1)	28.48	31.48	25.93	29.25	27.95
Income > $40,000 (% = 1)	10.13	9.26	12.59	9.91	10.95
Mother no high school (% = 1)	3.16	7.41*	8.15	4.25	5.76
Mother some college (% = 1)	50.00	31.48**	41.48	45.28	43.80
Mother college grad (% = 1)	24.05	29.63	18.52	25.47	22.77
N	158	54	135	212	347

ITT, Intention-to-Treat.
Decline group significantly different from choice group at ***$p < 0.000$; **$p < 0.01$; *$p < 0.10$.

among the students that appeared for study. This may be the case even if—as one would expect—there is a systematic difference between those who responded to the request for study and those who did not. Such a difference could contaminate the effect estimate only if treatment–control group differences varied systematically between the responders and nonresponders, and this variation directly influenced the voucher effect itself. The one pre-study measure available to gauge the respondent and nonrespondent groups in Charlotte suggests this is not the case. Of all original applicants to the Charlotte program, those who responded to the request for study participation, and those who did not, reported statistically significant income differences nine months earlier when the vouchers were awarded at random. Crucially, however, these differences did not vary between treatment and control groups. Greene (2000, 2001) reports that the respondent/nonrespondent income gap was roughly $3,000 for both groups.[6]

The second indication that the voucher effect need not be contaminated by nonresponse is that I determined the success of the initial randomization itself based only on the data drawn from the observable group of students who returned for study. Among the 347 unique student-level observations in the sample for which complete data exist, no demographic treatment–control differences are statistically significant at $p \leq 0.10$. Thus, although the randomization itself occurred nine months earlier, no measurable difference between those offered treatment and those in the control group *who self-selected into the study itself* is evident. The entire participant group, then, appears to be a statistically identical group with one random difference: the presence or absence of the voucher treatment. In this sense, the analytical sample represents a self-selected participant group—similar to those enticed to join a medical drug trial, for example—whose treatment condition is determined at random. As King et al. (2007) note, even in the absence of random selection of participants from the population, "valid causal inferences for the subjects of the experiment, although not necessarily for the population at large" (p. 486) can be produced if other factors, in particular that the treatment is a random condition among participants, hold within the study. Although other ideal conditions (large sample sizes, repeated effect measures) are not present here, a point estimate of the achievement difference between participating voucher students and public school students, at one moment in time, is obtainable.

Initial Analysis

This article concerns differences between three groups: students who were offered and accepted the voucher (choice), students who were offered but rejected the voucher (decline), and students who were not offered the voucher (control). As such, an initial comparison among characteristics of these groups of students is warranted here. As Table 1 indicates, students who declined the random voucher were statistically significantly more likely to be African American (although students in all subgroups are overwhelmingly black), less likely to live with both parents, more likely to have a mother who failed to complete high school, and less likely to have a mother who attended some college.

Table 2. ITT Estimates of the Voucher Effect

Parameter	Math		Reading	
	Estimate	s.e.	Estimate	s.e.
ITT	5.00	2.76*	6.14	2.82**
African American mother	−13.37	3.50***	−11.14	3.57***
Two parents	5.55	3.31*	12.37	3.38***
Male student	0.87	2.69	−6.11	2.75**
Public assistance	−5.42	3.13*	−9.02	3.20***
Income < \$20,000	−3.46	3.44	2.38	3.51
Income > \$40,000	0.11	4.66	−0.05	4.76
Mother no high school	−14.24	6.20***	−20.57	6.34***
Mother some college	3.17	3.26	0.93	3.33
Mother college grad	9.43	3.88**	13.55	3.96***
Intercept	37.61	4.81***	41.25	4.89***
N	347		347	
R^2	0.15		0.23	

ITT, Intention-to-Treat.
Notes: Income variables are compared to the omitted category of income between \$20,000 and \$40,000; mother's education variables are compared to the omitted category of students with mothers whose highest level of education was high school. Standard errors are robust ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$; two-tailed.

I begin the analysis of a voucher impact by estimating a typical "Intention-to-Treat" (ITT) model. In this model, students are considered to receive the treatment regardless of whether they used the voucher (ITT = 1 if assigned to voucher treatment, ITT = 0 if assigned to the control group).

This initial model is specified:

$$Y_i = \alpha + \beta\,\mathrm{ITT} + \gamma x_i + \varepsilon \tag{1}$$

where Y_i is student outcomes in math and reading, β is the impact of the voucher offer on student (math and reading) achievement, and x_i is a vector of control variables such as race, family structure, income, and parental education, with impact γ on achievement and model error ε. I report estimates of this model for math and reading scores in Table 2.

The ITT results indicate a positive voucher impact of 5 points on math scores and roughly 6 points on reading scores, all else equal. African American students are significantly disadvantaged on both tests relative to their peers, as are students without both parents in their home, and boys relative to girls. Students with less educated mothers are similarly disadvantaged, as are those from households that seek some form of public assistance (e.g., food stamps).

Next, I follow Howell and Peterson (2002) and Greene (2000) in estimating the mean effect of the voucher treatment using an IV analysis, where the instrument for the treatment is the random voucher offer itself (e.g., Heckman, 1996). Table 3 provides results. The results are similar to the ITT estimates in their statistical significance: A positive voucher effect appears evident at $p \leq 0.10$ for math achievement and $p \leq 0.05$ for reading. The point estimates of the voucher effect increase from 5 to nearly 7 points in math, and from 6 to 8 points in reading. This is not surprising since

Table 3. Instrumental Variable Estimates of the Voucher Effect

Parameter	Math		Reading	
	Estimate	s.e.	Estimate	s.e.
Treatment	6.85	3.77*	8.41	3.86**
African American mother	−12.79	3.51***	−10.44	3.59***
Two parents	4.47	3.39	11.04	3.47***
Male student	0.66	2.70	−6.37	2.76**
Public assistance	−5.44	3.13*	−9.05	3.20***
Income < \$20,000	−3.84	3.46	1.91	3.54
Income > \$40,000	0.63	4.69	0.59	4.79
Mother no high school	−13.71	6.23**	−19.91	6.37***
Mother some college	2.48	3.31	0.09	3.39
Mother college grad	9.29	3.88**	13.39	3.98***
Intercept	37.89	4.75***	41.60	4.86***
N	347		347	
R^2	0.15		0.23	

Notes: Income variables are compared to the omitted category of income between \$20,000 and \$40,000; mother's education variables are compared to the omitted category of students with mothers whose highest level of education was high school. Standard errors are robust ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$; two-tailed.

Table 4. Assigning Complier Values to Indicator Variables

	c_i	n_i
Compliers	1	0
Noncompliers	0	1

Notes: Compliers are defined as those offered treatment and accepting, or those assigned to control group. Noncompliers are defined as the "decline" group in previous sections: i.e., those assigned voucher treatment but declined it.

the ITT estimate is weighted down by students in the treatment group who declined the offer, and if a positive effect exists on average, these students were not exposed to the benefit. The IV estimates provide the most compelling evidence of a positive voucher effect in Charlotte. They are, moreover, similar in magnitude to previous research, particularly Howell and Peterson (2002). The central argument of this article, however, is that such evidence does not provide a full understanding of the policy effect of the voucher intervention.

CACE Estimation and Results

I now turn to the model on which this article is focused. Following Borman and Dowling (2006) and Jo (2002), I begin my estimation of the CACE by creating two indicator variables, whose values I assign as described in Table 4.

I then include these variables c_i and n_i in the following latent variable mixture model:

$$Y_i = \alpha_n n_i + \alpha_c c_i + \gamma_c c_i Z_i + \lambda_n n_i x_i + \lambda_c c_i x_i + \varepsilon_{ni} + \varepsilon_{ci} \quad (2)$$

where Y_i is continuous, and in this case either math or reading achievement test scores; α_n and α_c are intercepts for noncompliers and compliers, respectively; x is the same vector of covariates used above (e.g., race, mother's education, income); γ_c is the complier average causal effect; λ_n and λ_c are the covariate effects on achievement for noncompliers and compliers, respectively; and normally distributed errors ε_{ni} and ε_{ci}, which have mean 0 and variances σ_{nc}.[2]

Compliance, c_i is predicted using logistic regression, where

$$P(c_i = 1|x_i) = \pi_{ci} \tag{3}$$

$$P(ci = 0|xi) = 1 - \pi_{ci} = \pi_{ni} \tag{4}$$

$$\text{logit}(\pi_{ci}) = \beta_0 + \beta_1 x_i \tag{5}$$

where π_{ci} represents the probability for student *i* of compliance, π_{ni} is the probability of noncompliance, with logit intercept β_0 and β_1 representing the relationship between covariates x_i and the probability of compliance.

The CACE analysis in this article uses Mplus (Muthen & Muthen, 1998–2006) software, which employs the EM algorithm. The expectation step (E) computes the probability of compliance, which is modeled in the equations above, and the maximization step (M) calculates parameter estimates in the model specifying student achievement as a function of the voucher treatment and other control variables (Borman & Dowling, 2006; Muthen & Muthen 1998–2006). I specify CACE models of the effect of voucher treatment on math and reading achievement. These models control for the same student characteristics as those in the ITT models above, and these variables are likewise allowed to predict compliance status. Table 5 displays the results of parameter estimates.

The CACE for vouchers on math achievement is not statistically significant at conventional levels ($p = 0.14$, two-tailed), and the point estimate of the CACE falls below that of the ITT estimate. For reading, the CACE estimate is comparable to the IV and ITT estimates of the average treatment effects. Moreover, as in the analyses above, other demographic variables are clearly related to achievement, and these operate in expected directions. There are large and negative effects on achievement associated with being African American and having a mother who was a high school dropout. There is a disadvantage to being a male, at least on reading achievement. Conversely, there are strong positive effects for those whose mother graduated from college and, for reading, of living in a two-parent home.

Perhaps most importantly, the lower field of Table 5 provides some indication for the forces driving these results. Namely, noncompliance itself is strongly associated with parental education and the presence of two parents in the home. Students whose parents attained some college education and those who came from two-parent homes are considerably more likely to comply with whatever treatment they are assigned. African Americans, on the other hand, are less likely to comply.[7] In this formulation (and suggested in Figure 1), it is evident that compliance is partly

Table 5. Complier Average Causal Effect Results for Math and Reading Achievement

Parameter	Math		Reading	
	Estimate	s.e.	Estimate	s.e.
Complier class model				
Treatment	4.93	3.30	7.21	3.48**
African American mother	−12.9	3.92***	−11.45	3.80***
Two parents	3.91	3.12	11.88	3.31***
Male student	0.42	2.6	−6.83	2.63***
Public assistance	−5.02	3.14	−9.47	3.14***
Income < $20,000	−4.62	3.23	1.68	3.51
Income > $40,000	0.31	4.64	−0.78	4.58
Mother no high school	−12.62	4.96***	−19.62	4.61***
Mother some college	2.75	3.02	0.69	3.17
Mother college grad	9.84	3.93***	14.13	3.79***
Intercept (compliers)	39.39	4.67***	42.12	4.98***
Intercept (noncompliers)	37.87	5.65***	44.57	5.55***
Latent Class Model				
African American mother	1.04	0.59*	0.98	0.59*
Two parents	−1.68	0.52***	−1.64	0.52***
Male student	−0.01	0.34	0	0.34
Public assistance	−0.01	0.34	0.05	0.35
Income < $20,000	−0.53	0.44	−0.55	0.44
Income > $40,000	0.46	0.68	0.56	0.68
Mother no high school	0.97	0.92	0.89	0.93
Mother some college	−0.98	0.46**	−0.94	0.47**
Mother college grad	−0.21	0.48	−0.23	0.49
Intercept	−0.96	0.79	−0.97	0.8
N	347		347	

Notes: Latent Class Model predicts noncompliance. Income variables are compared to the omitted category of income between $20,000 and $40,000; mother's education variables are compared to the omitted category of students with mothers whose highest level of education was high school. Standard errors are robust ***$p < 0.01$; **$p < 0.05$; *$p < 0.10$; two-tailed.

determined by race, parental education, and family structure, and mediates the effects of private schooling on student achievement.

4. Discussion

The objective of this article is to argue that the CACE is the outcome of interest to researchers evaluating school voucher interventions. The CACE accounts for problems of self-selection within a voucher treatment group and explicitly accounts for the presence of voucher "always-takers" and "never-takers." It is this second advantage that should appeal in particular to policy analysts interested in the effectiveness of a targeted intervention: A CACE model provides information about which students will choose private school only when given the voucher opportunity to do so and which will not make such a choice, even if given the chance.

In the course of this argument, I have provided, as a substantive example, the results of CACE estimations using data from a small-scale field trial in Charlotte, NC. These results, which show positive effects of school vouchers for compliers, should be interpreted with some degree of caution. The data analyzed here are hardly ideal. The most obvious problem is that response rates for study participants were dangerously low. There was no pretest administered to students in any group, so no estimate of a change in scores may be obtained. Moreover, the trial lasted for only one year, and its results must be interpreted as a snapshot in time, now nearly eight years old.

It is important to note, however, that data on school choice programs—especially those involving vouchers or other types of scholarships—are still scarce. There have been only a handful of full-scale voucher interventions, and the leading studies of voucher outcomes have appropriately focused on data from these projects. I would argue that other data, however imperfect, may nevertheless be valuable to a literature in which evidence itself is a luxury.

Vouchers are inherently a matter of individual choice, even under ideal experimental conditions. Although attrition and noncompliance may be problems explained by a host of factors, school choice research necessarily encounters an active decision to remain in or break free from the status quo. Previous research has emphasized the existence of these factors in systematically partitioning the treatment group into those who accept and those who decline the voucher offer. However, I argue here that in evaluations of a voucher intervention, these factors are of substantive interest apart from the statistical need to account for them.

The evidence from this article should compel researchers on school vouchers to continually evaluate their methodological options. Compliance mediates the impact of a voucher intervention and, for math achievement, the estimated CACE is not statistically significant. Because compliance appears related, at minimum, to students' race, parental education, and family structure (here the presence of two parents in the home), this difference may not be trivial. Such characteristics are certainly of interest to policymakers concerned with education—especially the education of high-risk children.

Joshua M. Cowen is a doctoral candidate in political science at the University of Wisconsin-Madison.

Notes

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant 144-NL14 to the University of Wisconsin-Madison. The opinions expressed are those of the author and do not represent views of the U.S. Department of Education. The author thanks Geoffrey Borman, David Kaplan, John Witte, and Patrick Wolf for helpful advice at various stages of the article, and Jay Greene for use of the data. All errors are the author's responsibility. Data and software syntax (Stata and Mplus) are available from the author by request.

1. Institute for Justice: National School Choice Timeline, May 2005.
2. See Witte (2000, chap. 4).
3. $33,000 compared to $30,700. See Howell and Peterson (2002, p. 72 and chap. 3) for more discussion.
4. In the economics literature this parameter is known as the Local Average Treatment Effect (LATE).
5. See, for example, Clogg (1995) for an introduction to latent class analysis.
6. Precisely, "lottery winners who participated in the study had average family incomes of $25,323, while lottery winners who did not participate had average family incomes of $22,517. Control group students

who participated in the study had average family incomes of $25,297, while control group students who did not participate had average family incomes of $22,215" (Greene, 2000).

7. Previous research (e.g., Howell & Peterson, 2002) has found voucher effects themselves to be limited to African Americans. The results of analyses presented here did not change significantly when limited only to African American participants.

References

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Barnard, John, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. Rubin. 2003. "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City." *Journal of the American Statistical Association* 98 (462): 299–311.

Bloom, Howard S. 1984. "Accounting for No-Shows in Experimental Evaluation." *Evaluation Review* 8 (2): 225–46.

Borman, Geoffrey D., and N. Maritza Dowling. 2006. "Longitudinal Achievement Effects of Multiyear Summer School: Evidence from the Teach Baltimore Randomized Field Trials." *Educational Evaluation and Policy Analysis* 28 (1): 25–48.

Clogg, Clifford C. 1995. "Latent Class Models." In *Handbook of Statistical Modeling in the Social and Behavior Sciences*, ed. Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. San Francisco: Jossey-Bass, 311–59.

Dunn, Graham, Mohommad Maracy, Christopher Dowrick, Jose Luis Ayuso-Mateos, Odd Steffen Dalgard, Helen Page, Ville Lehtinen, Patricia Casey, Clare Wilkinson, Jose Luis Vazques-Barquero, and Greg Wilkinson. 2003. "Estimating Psychological Treatment Effects from a Randomized Controlled Trial with Both Noncompliance and Loss to Follow-Up." *British Journal of Psychiatry* 323–31.

Greene, Jay P. 2000. "The Effect of School Choice: An Evaluation of the Charlotte Children's Scholarship Fund." Manhattan Institute Civic Report, New York.

———. 2001. "The Effect of School Choice: An Evaluation of the Charlotte Children's Scholarship Fund." *Education Next* 1 (2): 55–60.

Greene, Jay P., Paul E. Peterson, and Jiangtau Du. 1999. "Effectiveness of School Choice: The Milwaukee Experiment." *Education and Urban Society* 31 (2): 190–213.

Have, Thomas R. Ten, Michael R. Elliott, Marshall Joffee, Elaine Zanutto, and Catherine Datto. 2004. "Causal Models for Randomized Physician Encouragement Trials in Treating Primary Care Depression." *Journal of the American Statistical Association* 99 (465): 16–25.

Heckman, James J. 1996. "Randomization as an Instrumental Variable." *Review of Economics and Statistics* 78 (2): 336–41.

———. 1997. "Instrumental Variables: A Study of Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32 (3): 441–62.

Howell, William G., and Paul E. Peterson (with Patrick J. Wolf and David E. Campbell) 2002. *The Education Gap: Vouchers and Urban Schools*. Washington, DC: Brookings Institution Press.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Imbens, Guido W., and Donald B. Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies* 64: 555–74.

Jo, Booil. 2002. "Model Misspecification Sensitivity Analysis in Estimating Causal Effects of Interventions with Noncompliance." *Statistics in Medicine* 21: 3161–81.

Jo, Booil, and Bengt O. Muthen. 2001. "Modeling of Intervention Effects with Noncompliance: A Latent Variable Approach for Randomized Trials." In *New Developments and Techniques in Structural Equation Modeling*, ed. George A. Marcoulides and Randall E. Schumacker. Mahwah, NJ: Lawrence Erlbaum Associates, 57–87.

King, Gary, Emmanuela Gakidou, Nirmala Ravishankar, Ryan T. Moore, Jason Lakin, Manett Vargas, Martha Maria Tellez-Rojo, Juan Eugenio Hernandez Avila, Mauricio Hernandez Avila, and Hector Hernandez Llamas. 2007. "A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal Health Insurance Program." *Journal of Policy Analysis and Management* 26 (3): 479–506.

Krueger, Alan B., and Pei Zhu. 2004a. "Another Look at the New York City Voucher Experiment." *American Behavioral Scientist* 47 (5): 658–98.

———. 2004b. "Inefficiency, Subsample Selection Bias, and Non-Robustness: A Response to Paul E. Peterson and William G. Howell." *American Behavioral Scientist* 47 (5): 718–28.

Little, Roderick J., and Donald B. Rubin. 1987. *Statistical Analysis with Missing Data.* New York: John Wiley.

Little, Roderick J., and Linda H. Y. Yau. 1998. "Statistical Techniques for Analyzing Data from Prevention Trials: Treatment of No-Shows Using Rubin's Causal Model." *Psychological Methods* 3 (2): 147–59.

Mealli, Fabrizia, Guido W. Imbens, Salvatore Ferro, and Biggeri Annibale. 2004. "Analyzing a Randomized Trial on Breast Self-Examination with Noncompliance and Missing Outcomes." *Biostatistics* 5: 207–22.

Muthen, Bengt O. 2001. "Second Generation Structural Equation Modeling with a Combination of Categorical and Continuous Latent Variables: New Opportunities for Latent Class–Latent Growth Modeling." In *New Methods for the Analysis of Change*, ed. Linda Collins and Aline Sayer. Washington, DC: American Psychological Association, 291–322.

Muthen, Linda K., and Bengt O. Muthen. 1998–2006. *Mplus (Version 4.1) Users Guide*, 4th ed. Los Angeles: Muthen & Muthen.

Peterson, Paul E., and William G. Howell. 2004. "Efficiency, Bias and Classification Schemes: A Response to Alan B. Krueger and Pei Zhu." *American Behavioral Scientist* 47 (5): 699–717.

Rouse, Cecilia Elena. 1998. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *The Quarterly Journal of Economics* 113 (2): 553–602.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688–701.

Witte, John F. 2000. *The Market Approach to Education: An Analysis of America's First Voucher Program.* Princeton, NJ: Princeton University Press.