



REVIEW OF *THE EFFECTS OF SCHOOL VOUCHERS ON COLLEGE ENROLLMENT*

Reviewed By

Sara Goldrick-Rab

University of Wisconsin-Madison

September 2012

Summary of Review

This Brookings report examines college enrollment rates of students participating in an experimental New York School Choice Scholarships Foundation Program, which in the spring of 1997 offered 3-year scholarships worth up to \$1,400 annually to low-income families. The study identifies no overall impacts of the voucher offer, but the authors report and emphasize large positive impacts for African American students, including increases in college attendance, full-time enrollment, and attendance at private, selective institutions of higher education. This strong focus on positive impacts for a single subgroup of students is not warranted. There are no statistically significant differences in the estimated impact for African Americans as compared to other students; there is important but unmentioned measurement error in the dependent variables (college attendance outcomes) affecting the precision of those estimates and likely moving at least some of them out of the realm of statistical significance; the authors fail to demonstrate any estimated negative effects that could help explain the average null results; and there are previously existing differences between the African American treatment and control groups on factors known to matter for college attendance (e.g., parental education). Contrary to the report's claim, the evidence presented suggests that in this New York City program, school vouchers did not improve college enrollment rates among all students or even among a selected subgroup of students.

Kevin Welner

Project Director

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado

Boulder, CO 80309-0249

Telephone: (802) 383-0058

Email: NEPC@colorado.edu

<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

REVIEW OF *THE EFFECTS OF SCHOOL VOUCHERS ON COLLEGE ENROLLMENT: EXPERIMENTAL EVIDENCE FROM NEW YORK CITY*

Sara Goldrick-Rab, University of Wisconsin-Madison

I. Introduction

Confronted with evidence that the United States lags behind its peers in rates of college attainment, educators and policymakers are eager to find ways to boost enrollment in postsecondary education, especially among students from disadvantaged families.¹ Research consistently demonstrates that k-12 schooling plays a critical role in setting students up for success or failure during the transition to college.² Matriculation to higher education, especially into selective colleges and universities, is particularly uncommon among children from low-income families enrolled in public urban schools.

But little seems to work when it comes to increasing participation in higher education. In particular, relatively few interventions appear to effectively expand college opportunities for low-income children of color. This new report from the Brookings Institution's Brown Center on Education Policy and Harvard University's Program on Education Policy and Governance, "The Effects of School Vouchers on College Enrollment: Experimental Evidence from New York City," is notable for purporting to identify a strategy that *does* work: providing students with vouchers to attend private schools. Authors Matthew M. Chingos and Paul E. Peterson contend that on average a privately funded New York City voucher program distributing scholarships to students beginning in 1997 exerted no overall improvements in college attendance, but instead boosted rates of college enrollment among African Americans by 24%.³

II. Findings and Conclusions of the Report

The report estimates the long-term impacts of the New York School Choice Scholarships Foundation Program on an established array of college indicators: on-time enrollment (within 3 years of high school graduation); full-time enrollment (which reduces time-to-degree); and location of enrollment (2-year/4-year sector, public/private control, and selectivity of the institution attended).

While the authors identify no average impacts of either the offer or use of the vouchers, they emphasize one finding: an estimated positive impact of vouchers for the African American students, who comprised 42% of the total sample of 2,642 students. They describe the magnitude of the impact as large and substantively important, noting that among African American students who received the voucher, college enrollment rates were 8.7 percentage points higher when compared to the 36% college enrollment rate of the African American control group not offered the grant. Some of this increase occurred via a statistically significant boost in rates of attendance at selective colleges and universities.

The report concludes that the voucher had positive impacts for African Americans but not for Hispanics because of unobserved advantages held by the latter group, which led them to have higher college attendance rates irrespective of the vouchers (p. 17). Among the observed advantages, Hispanic parents rated the quality of their children's schools as of great importance, while African American parents rated the impact of the voucher itself as more important. The authors also suggest that the effects of the voucher might differ because families used them differently—with Hispanics more apt to choose a religious school, while African Americans optimized their “secular educational objectives” (p. 18). Regardless of the reason, the reader is told that the effects of vouchers on college attendance are “unusually large” and that the voucher “has a much larger impact than does exposure to an effective teacher” (p. 20). In the popular press, the authors have contended that the clear implication and necessary next step is for President Obama to promote college attainment among African American students by opening “private-school doors for low-income students.”⁴

III. The Report's Rationale for Its Findings and Conclusions

The findings are based on estimates from a randomized trial of a New York City program that distributed vouchers up to \$1,400 to students from low-income families beginning in spring 1997. The experiment and its results have been discussed in many other volumes and articles, and they have been subject to critical debate.⁵ The new results come from a longitudinal follow-up in which student records were traced using the National Student Clearinghouse, providing an opportunity to assess impacts on college enrollment between 1997 and 2011. The authors estimated impacts of the voucher (both offer and receipt, which had similar impacts) on average, and for two selected racial/ethnic subgroups: African Americans and Hispanics.

IV. The Report's Use of Research Literature

The report leverages prior research literature to make three points: (1) some interventions have few short-term impacts, but instead have longer-lasting results only revealed by follow-up studies; (2) there is reason to think that longer-term effects might emerge for

interventions affecting school quality (like vouchers); and (3) prior evidence suggests heterogeneous impacts from school vouchers. There is little reason to question the first two points. Regarding the third, the report downplays the controversy surrounding that prior evidence. Pointing to only those studies that support the authors' claims of the validity of those effects,⁶ and failing to make any mention of the studies questioning those estimations,⁷ the report leaves a less-informed reader without important knowledge of that debate. For example, discussion of a key issue—how students are classified by their race/ethnicity for analysis—is relegated to the sixteenth footnote, and while the authors mention a critique of their approach, they provide just one single counter-argument and offer no additional testing to support that argument.

V. Review of the Report's Methods

It is clear that the report's authors endeavored to respond to past critiques about methodology in studies of this New York City program. For example, they included all students in the analysis (rather than only those with baseline test score data),⁸ and used a data source for measuring outcomes that does not have much missing data (but still has measurement error; see below). They also limited the assessment of impacts to a set of specific outcomes, minimizing the chances that they would find a significant result simply by chance alone.

Moreover, there is certainly much to appreciate in the authors' consideration of the potential for heterogeneous effects. The research community is interested in the possibility that treatments work differently for different groups of students, and this has led to studies that investigate heterogeneity rigorously and appropriately, so as to minimize the risk of false positives. For example, careful analyses of the Tennessee STAR dataset have suggested that smaller class sizes in early years yield relatively greater benefits for children who are disadvantaged.⁹ In exploring the possibility of heterogeneous effects, the best studies avoid the use of *ad hoc* subgroup analyses. Instead, they ground modeling strategies in strong theories and identify treatment effect heterogeneity in a statistically principled manner. Such work also conveys that subgroup analyses are exploratory approaches to building theory and future hypotheses for more rigorous testing—not necessarily confirmatory evaluations on which public policy should be based.¹⁰

While the report's estimation of average treatment impacts appears convincing, effect heterogeneity is not well established in the analysis. In part, this is because the modeling strategies are not grounded in strong theories. But the more important problem lies in the authors' failure to rule out the possibility that there are no differences between the African American and Hispanic subgroups.

When testing for heterogeneous impacts, researchers should seek to rule out the possibility that there are *no subgroup differences*. Typically researchers test this hypothesis by estimating a subgroup-by-treatment interaction term and displaying impacts for each subgroup along with an indication of whether those impacts differ from one another.

“Findings for a specific subgroup should not be highlighted unless they differ statistically significantly from those for other sample members. If subgroup differences are not statistically significant, findings for the full study sample usually should be emphasized instead of those for the subgroup.”¹¹ That has not been the practice with regard to analyses of this New York City intervention; in 2000, Mathematica Policy Research issued a similar warning: “Because gains are so concentrated in this single group, one needs to be very cautious.”¹²

But in this report there is *no consistent indication of whether the estimated impacts for African Americans and Hispanics are statistically different from one another*. Twice in the text, the authors state that they are not statistically different from one another. But when displaying impact estimates in Tables 3, 4, and 5, the report shows results for the full sample, and then selectively displays only the African American and Hispanic subsamples without anything in the table showing whether their results are truly different. A useful modification would have been to include, at the bottom of the panel for each subgroup, an F-statistic and p-value for testing the ‘null hypothesis’ that the subgroup means are equal. Were this provided, the reader would observe that the authors *cannot* reject the equality of the treatment effects by race/ethnicity. This is a crucial point: despite appearances—one column (African Americans) has “stars,” which normally indicate statistical significance, and the other (Hispanics) does not—they are not different from one another.¹³

In addition, there is measurement error in the dependent variables that likely affects the subgroup analyses but is not mentioned in the report. It is clear that the authors worked hard to attend to the intricacies of the National Student Clearinghouse (NSC) data.¹⁴ But their claims of the quality of the NSC data are overstated. First, they failed to note that while the NSC records college enrollment at 96% of the nation’s colleges and universities, the coverage rate varies—it is strongest at publics, and weaker at private colleges and for-profit colleges.¹⁵ Second, the NSC’s capacity to accurately determine college enrollment for a student depends on the reliability of the fuzzy matching algorithm it employs. The report claims “of the 2,666 students in the original study, the information needed to match the data was available for 2,642, or 99.1% of the original sample” (p.iii), emphasizing the availability of students’ social security numbers provided in the file to NSC for the matching process. But the NSC *did not use* SSNs to make the match—the matches were made based solely on name and date of birth, a process that may be more fallible for students with more complexity to their names, for students with very common names, for student groups more likely to have missing data (e.g., racial/ethnic minority students), or both.¹⁶ Thus, the report oversells the degree to which using the NSC solves the attrition problem confronted in prior studies of the program. The measurement error in the dependent variable, even if it is not correlated with treatment, affects the size of the standard errors. Taking for example the key findings about positive impacts on attendance at private and selective institutions for African Americans (Table 6), if we inflate the standard errors even slightly, the estimates will be rendered statistically non-significant. Once again, this implies there is likely no story in the subgroups.

Furthermore, readers are told nothing about the impact estimates for the students who are not in either of those groups. In the popular press, the authors contend that this sample of

students was too small for analysis.¹⁷ But providing information about the estimated impact on these students would inform readers. Might it have been negative? Given a null average treatment effect and a positive impact for 42% of the students, only a substantial negative for that 12% of the sample excluded from the presented analyses would make sense. But no such difference is reported.

The exclusion of results for students who are neither African American nor Hispanic is also odd given that the approach diverges from Professor Peterson's earlier work, in which subgroup estimates are shown for African Americans versus all other students.¹⁸ In this report the reader is simply told that the reported "105" white and Asian students appear are *not equivalent* on observable characteristics at baseline and are therefore omitted from the sample (p. 10). The "105" is in quotation marks since this is the number in the text.

The most precise estimate in this report does not provide evidence that the vouchers were effective in advancing the participation of students in higher education.

Simple subtraction using the numbers presented in the tables suggests there were actually 323 missing students (subtract all African American and Hispanic students from the total number of students). The reason for the discrepancy may relate to problems of racial classification, but this is not made clear. In a prior study of the same program by Peterson and his colleagues, a re-analysis found that estimated positive impacts for African Americans were rendered null when children with an African American non-Hispanic *father* were classified as African Americans along with those born to an African American mother.¹⁹ Might that be the case in this study as well?

The report also downplays another challenge to the validity of the results for African Americans—the baseline differences in levels of parental education for that subgroup. On page 10, readers are assured that the treatment and control groups are similar on average, and that statistical testing indicates no problems for the two subgroups shown (African Americans and Hispanics). But in this analysis there is non-equivalence of parental education. Since parental education has well-established explanatory power for the dependent variable—college enrollment—this difference is critical. The treatment group of African Americans has a higher rate of parental bachelor's degree completion than the control group. That we cannot statistically reject selection on this crucial observable characteristic certainly raises reasonable concerns that selection on unobservable variables is also operating.²⁰ A stronger analysis would have done more to estimate the degree to which omitted variables bias affects the estimates of these subgroup impacts.²¹ That is, readers cannot know whether any positive college enrollment differences for African Americans are due to vouchers or due to pre-existing differences among students—such as whether their parents completed a bachelor's degree.

Finally, in reasoning *why* African Americans might have benefited from vouchers while Hispanics did not, the authors suggest that unobservable characteristics are to blame, but

they are vague about which types of important factors are unobserved. They do emphasize differences in two observable variables (perceived school quality and perceived voucher impact) but these are two of many possible measures. They also posit a potential mechanism—racial differences in the types of schools that students with vouchers chose to attend (religious versus other)—but make no effort to examine that causal claim. For example, is there any evidence that the voucher increased college enrollment only for students who chose non-religious private schools? The authors do little more than nod at potential mechanisms, but explanations of this sort are arguably critical when an analysis ends up focusing on heterogeneity—on treatment effects for some students but not for others. This is unfortunate since the failure to explore alternate hypotheses that are falsifiable can lead to confirmatory bias. When estimated findings are particularly important or surprising, it is especially critical to consider such alternative explanations. In this report, only the case for the veracity of the subgroup differences is laid out—despite the fact, as noted above, that the subgroup differences are not statistically significant.

VI. Review of the Validity of the Findings and Conclusions

Contrary to how it was presented, the main finding of this new report *should* be that, using a rigorous experimental design in which vouchers were randomly assigned to students, the estimated college enrollment rates of students with and without vouchers *were not different* from one another. In other words, the most precise estimate in this report does not provide evidence that the vouchers were effective in advancing the participation of students in higher education.

Moreover, and again contrary to how it was presented, there is no story in the subgroups. The estimated effects for the two racial/ethnic subgroups are not statistically significantly different from each other, and those estimates are subject to important measurement error that was not mentioned in the report. Hence, there is not strong evidence that the statistically significant result for African Americans that is set forth in the report is truly statistically significant or different from the non-statistically significant result for Hispanics. It is the full sample finding, showing no effects of vouchers, which deserves the most attention and merits policy scrutiny.

Despite this, the main finding of null average treatment effects receives just one paragraph of text (on page 12) and instead the subgroup results are emphasized throughout the report. The authors devote five full pages to convincing the reader that there were positive impacts of vouchers for African Americans, and they present this as the *only* finding worthy of elaboration. Potential explanations for the null result are never explored, and whenever the null findings are reported, the impacts for the African American subgroup are jointly reported either in the same sentence or immediately following it (for example, see pp. ii, iii, and 5). In light of the overall null impact of vouchers, this would suggest the presence of a substantial negative effect for students who are non-Hispanic white or Asian—which is a result never mentioned in the report.

VII. Usefulness of the Report for Guidance of Policy and Practice

Policymakers and practitioners interested in the effectiveness of school voucher programs should indeed attend to the results of this study, which—contrary to the interpretation of the authors—convincingly demonstrates that in New York City a private voucher program failed to increase the college enrollment rates of students from low-income families.

Notes and References

1 See <http://www.ed.gov/college-completion> .

Russell, A. (2011). *A Guide to Major College Completion Agendas*. Washington, DC: American Association of State Colleges and Universities. Retrieved September 7, 2012, from <http://www.aascu.org/policy/publications/policymatters/2011/collegecompletion.pdf>.

Goldrick-Rab, S. (2010). Challenges and opportunities for improving community college student outcomes. *Review of Educational Research*, 80(3), 437-469.

2 Adelman, C. (2006). *The Toolbox Revisited: Paths to Degree Completion From High School Through College*. Washington, DC: U.S. Department of Education. Retrieved September 7, 2012, from <http://www.ed.gov/rschstat/research/pubs/toolboxrevisit/index.html>.

Roderick, M. et al. (2008). *From High School to the Future: Potholes on the Road to College*. Chicago: Consortium for Chicago School Research. Retrieved September 7, 2012, from http://ccsr.uchicago.edu/downloads/1835ccsr_potholes_summary.pdf.

3 Chingos, M. M. & Peterson, Paul E. (2012). *The Effects of School Vouchers on College Enrollment: Experimental Evidence from New York City*. Washington, DC: The Brookings Institution. Retrieved September 7, 2012, from <http://www.brookings.edu/research/papers/2012/08/23-school-vouchers-harvard-chingos>.

4 Chingos, M.M. and Peterson, P. E. (2012, August 23). A generation of school-voucher success. *Wall Street Journal*. Retrieved September 7, 2012, from <http://online.wsj.com/article/SB10000872396390444184704577585582150808386.html> (subscription required).

5 Howell, W.G. & Peterson, P.E. (2006). *The Education Gap*. Washington, DC: The Brookings Institution.

6 Howell, W.G. & Peterson, P.E. (2004). Uses of theory in randomized field trials: lessons from school voucher research on disaggregation, missing data, and the generalization of findings. *American Behavioral Scientist*, 47 (5), 634-657.

Mayer, D.P., Peterson, P.E., Myers, D.E., Tuttle, C.C., & Howell, W.G. (2002). *School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program: Final Report*, MPR, 8404-045. Princeton and Cambridge: Mathematica Policy Research and the Program on Education Policy and Governance at Harvard University.

7 Krueger, A. & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, 47 (5), 658-698.

8 The rationale is that there is no need to control for baseline scores when using randomized assignment to a control group. See:

Krueger, A. & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, 47 (5), 658-698.

9 Krueger, A.B. & Whitmore, D.M. (2001). *Would Smaller Classes Help Close the Black-White Achievement Gap?* Princeton, NJ: Princeton University.

10 Peterson has argued that analysis of subgroups is an important way in which theory should be used to guide empirical research (cf. Howell & Peterson 2004). I concur with this and have argued as much in my own work; for example, see Harris, D.N. & Goldrick-Rab, S. (2012). But as Peterson notes, “When selecting among comparisons, however, the analyst ultimately needs theory, because rampant empiricism cannot distinguish idiosyncratic from genuine findings” (Howell and Peterson, 2004, p. 642). More than theory is necessary, according to methodologists like Bloom and Michaelopoulos (2011), who note that post-hoc subgroup analyses should be described as exploratory and suggestive, rather than confirmatory, and should be subjected to all of the useful rigorous statistical tests. See:

Howell, W.G. & Peterson, P.E. (2004). Uses of theory in randomized field trials: lessons from school voucher research on disaggregation, missing data, and the generalization of findings. *American Behavioral Scientist*, 47 (5), 634-657;

Harris, D.N. & Goldrick-Rab, S. (2012) Improving the productivity of education experiments: Lessons from a randomized study of need-based financial aid. *Educational Finance and Policy*, 7(2), 143-169;

Bloom, H. & Michaelopoulos, C. (2011, January) When is the story in the subgroups? : Strategies for interpreting and reporting intervention effects for subgroups. *Prevention Science* (online).
<http://dx.doi.org/10.1007/s11121-010-0198-x> (subscription or payment required).
A draft is archived at <http://www.mdrc.org/publications/551/full.pdf>.

11 Bloom, H. & Michaelopoulos, C. (2011, January) When is the story in the subgroups? : Strategies for interpreting and reporting intervention effects for subgroups. *Prevention Science*. Retrieved September 7, 2012, from <http://www.mdrc.org/publications/551/full.pdf>, 5.

12 Mathematica Policy Research, Inc. (2000). Voucher claims of success are premature in New York City (press release). Princeton, NJ: Author. (No longer available online.)

13 The same approach is employed in Mayer et al (2002). On p. 35 in that report, authors note that sample sizes for subgroups are small, do not test for significant differences for them, and yet repeatedly emphasize the effects for African Americans.

Mayer, D.P., Peterson, P.E., Myers, D.E., Tuttle, C.C., & Howell, W.G. (2002). *School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program: Final Report*, MPR, 8404-045. Princeton and Cambridge: Mathematica Policy Research and the Program on Education Policy and Governance at Harvard University.

14 Goldrick-Rab, S. & Harris, D. (2010). Observations on the use of NSC data for research purposes (White paper). Chicago: The Spencer Foundation (<http://www.spencer.org>). Retrieved September 7, 2012, from <http://www.finaidstudy.org/documents/NSC%20Dear%20colleagues%20letter.pdf>.

15 Dynarski, S. (2012, September 5). Personal communication with the author regarding Dynarski’s forthcoming paper.

16 After a reminder that NSC does not utilize SSNs for matching for research studies, author Chingos double-checked his assumption and informed me that SSN was not used.

17 Chingos, M.M. and Peterson, P. E. (2012, August 23). A generation of school-voucher success. *Wall Street Journal*. Retrieved September 7, 2012, from <http://online.wsj.com/article/SB10000872396390444184704577585582150808386.html> (subscription required).

18 Howell, W.G. & Peterson, P.E. (2004). Uses of theory in randomized field trials: lessons from school voucher research on disaggregation, missing data, and the generalization of findings. *American Behavioral Scientist*, 47 (5), 634-657.

19 Krueger, A. & Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist*, 47 (5), 658-698.

20 While other imbalanced observables point in the opposite direction, possibly anticipating a negative impact, those observables bear a weaker relationship to the outcome of interest.

21 For example, see:

Altonji, J., Elder, T., & Taber, C. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1), 151-184.

DOCUMENT REVIEWED:

**The Effects of School Vouchers on
College Enrollment: Experimental
Evidence from New York City**

AUTHORS:

Matthew M. Chingos and Paul E. Peterson

PUBLISHER/THINK TANKS:

Brown Center on Education Policy at
Brookings and Harvard Kennedy Program on
Education Policy and Governance

DOCUMENT RELEASE DATE:

August 23, 2012

REVIEW DATE:

September 13, 2012

REVIEWER:

Sara Goldrick-Rab, University of Wisconsin-
Madison

E-MAIL ADDRESS:

srab@education.wisc.edu

PHONE NUMBER:

(608) 265-2141

SUGGESTED CITATION:

Goldrick-Rab, S. (2012). *Review of "The Effects of School Vouchers on College Enrollment: Experimental Evidence from New York City."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-vouchers-college>.