

5-1-2008

School Voucher Programs: What the Research Says About Parental School Choice

Patrick J. Wolf

Follow this and additional works at: <https://digitalcommons.law.byu.edu/lawreview>



Part of the [Education Law Commons](#)

Recommended Citation

Patrick J. Wolf, *School Voucher Programs: What the Research Says About Parental School Choice*, 2008 BYU L. Rev. 415 (2008).

Available at: <https://digitalcommons.law.byu.edu/lawreview/vol2008/iss2/9>

This Article is brought to you for free and open access by the Brigham Young University Law Review at BYU Law Digital Commons. It has been accepted for inclusion in BYU Law Review by an authorized editor of BYU Law Digital Commons. For more information, please contact hunterlawlibrary@byu.edu.

School Voucher Programs: What the Research Says About Parental School Choice

Patrick J. Wolf

I. INTRODUCTION

A number of important policy questions surround school voucher initiatives. Before a new voucher program is enacted, policymakers usually want to know answers to questions such as: (1) Do voucher programs primarily serve disadvantaged students?; (2) Do parents like voucher programs?; and (3) Do students benefit academically from vouchers? The answers to these questions provide policymakers and the general public with crucial information regarding what societal goals are and are not advanced when parents are allowed to use public funds to enroll their child in a private school of their choosing.

Fortunately, enough voucher programs have been established and evaluated to provide us with consistent and reliable answers to many of the policy questions surrounding school vouchers targeted at disadvantaged students. Had the Utah universal school voucher program not been defeated in a recent public referendum, it would have been the thirteenth school voucher program launched in the United States.¹ The Utah initiative would have been the first voucher program in this country open to all school-age children.² The twelve

* Professor and Twenty-first Century Chair in School Choice, Department of Education Reform, College of Education and Health Professions, University of Arkansas.

1 See *infra* Table 1. See generally MILTON & ROSE D. FRIEDMAN FOUNDATION, THE ABCS OF SCHOOL CHOICE (2006–2007), available at <http://www.friedmanfoundation.org/friedman/downloadFile.do?id=102> [hereinafter FRIEDMAN FOUNDATION].

2. See FRIEDMAN FOUNDATION, *supra* note 1, at 46–47. Chile has operated a universal school voucher program since 1982. For information on that program, see Claudio Sapelli, *The Chilean Education Voucher System*, in WHAT AMERICA CAN LEARN FROM SCHOOL CHOICE IN OTHER COUNTRIES 41, 41–62 (David Salisbury & James Tooley eds., 2005). European and Commonwealth countries such as the Netherlands, Belgium, and several Canadian provinces provide full government funding to students who attend qualified private secular and religious schools. In the Netherlands, for example, nearly seventy percent of K–12 students attend religious schools at public expense. Such school choice arrangements are distinct from American-style school vouchers in that the payments are made directly to the schools, not to

voucher programs that are approved and operating in the United States³ all target voucher eligibility to students that are disadvantaged in various ways. Thus, the research to date on school vouchers provides only speculative information about the likely effects of universal programs even as it provides a wealth of data on the effects of the targeted voucher programs that are becoming an increasingly common feature of the school-reform landscape.

The high-quality studies on school voucher programs generally reach positive conclusions about vouchers. The many evaluations of targeted school voucher initiatives confirm that these programs serve highly disadvantaged populations of students. Of the ten separate analyses of data from “gold standard” experimental studies of voucher programs, nine conclude that some or all of the participants benefited academically from using a voucher to attend a private school. The evidence to date suggests that school voucher programs benefit many of the disadvantaged students and parents that they serve.⁴

Part II of this Article describes the twelve voucher programs that currently exist in the United States and the student populations that they serve. Part III discusses and critiques the various methods that have been used to evaluate school voucher programs. Part IV argues that the evidence from rigorous voucher evaluations indicates that voucher programs increase parental satisfaction with schools and tend to boost student test scores, at least for some participants.

the parents, and the money comes with extensive regulatory strings attached. Stephen Macedo & Patrick J. Wolf, *Introduction: School Choice, Civic Values, and Problems of Policy Comparison*, in *EDUCATING CITIZENS* 1, 9–12, 15–17 (Patrick J. Wolf & Stephen Macedo eds., 2004).

3. As our interest here is in the voucher programs and evaluations in the United States, the reader should understand the domain under discussion to be limited to the United States unless otherwise specified.

4. Due to space limitations, this Article does not discuss the potential “systemic” or competitive effects of school voucher programs on public schools and the students that attend them. Many empirical studies speak to that question. Interested readers should see Clive R. Belfield & Henry M. Levin, *The Effects of Competition Between Schools on Educational Outcomes: A Review for the United States*, 72 *REV. EDUC. RES.* 279 (2002), and Caroline Minter Hoxby, *Rising Tide*, *EDUC. NEXT*, Winter 2001, at 68, for reviews of the theory and evidence regarding the systemic effects of school vouchers. The Article also does not consider the effects of voucher programs on civic values and the public purposes of education. Although that is a crucial concern, that topic is effectively explored in David Campbell’s article in this volume as well as two previous articles of mine. See Patrick J. Wolf, *Civics Exam: Schools of Choice Boost Civic Values*, *EDUC. NEXT*, Summer 2007, at 66; Patrick J. Wolf, *School Choice and Civic Values*, in *GETTING CHOICE RIGHT: INSURING EQUITY AND EFFICIENCY IN EDUCATION POLICY* (Julian R. Betts & Tom Loveless eds., 2004).

Readers are cautioned that this evidence is drawn from targeted voucher programs and may not apply to universal programs such as the one proposed for and subsequently rejected by the citizens of Utah. Part V concludes by encouraging more rigorous research on the impacts of voucher programs with various design features.

II. SCHOOL VOUCHER PROGRAMS IN THE UNITED STATES

A school voucher program is an arrangement whereby public funds are made available to qualified parents to cover some or all of the expenses associated with enrolling their child in a participating private school of their choosing. Privately funded scholarships are not school vouchers, although, like vouchers, they are used to allow disadvantaged students to gain access to private schools. The placement and funding of special needs students in private schools by public school districts also is not a voucher program, since district officials, and not parents, choose the school. The definitional aspects of school vouchers are the source of the funds (governmental), the purpose for which the funds are provided (to enroll a school-age child in a private school), and the party whose decisions fulfill that purpose (a parent or legal guardian of the child).⁵

According to this definition of school vouchers, twelve voucher programs had been established or were being implemented in the United States as of the fall of 2007.⁶ A total of 56,285 students were enrolled in these programs at the start of the 2006–2007 school year.⁷ America's first school voucher program was established in Vermont in 1869. The Vermont "town tuitioning" program provides vouchers for students in rural areas without public junior high or high schools.⁸ The vouchers in most towns enable parents to enroll their children in the public or private high school of their

5. Patrick J. Wolf, *Vouchers*, in THE ROUTLEDGE INTERNATIONAL ENCYCLOPEDIA OF EDUCATION 635, 635 (Gary McCulloch & David Crook eds., 2008).

6. See *infra* Table 1.

7. *Id.* Thousands of students displaced by Hurricanes Katrina or Rita received federally funded school vouchers from 2005 to 2007. Because these emergency school vouchers were temporary, those students are not included in this count. See Friedman Foundation Newsroom, The School Choice Advocate, <http://www.friedmanfoundation.org/friedman/newsroom/ShowArticle.do?id=23> (last visited April 4, 2008).

8. FRIEDMAN FOUNDATION *supra* note 1, at 48; Christopher W. Hammons, *The Effects of Town Tuitioning in Vermont and Maine*, 1 SCHOOL CHOICE ISSUES IN DEPTH 5 (Milton & Rose D. Friedman Foundation 2001), available at <http://www.friedmanfoundation.org/friedman/downloadFile.do?id=61>.

choosing.⁹ Other towns send all their students to one school. A similar program has operated in Maine since 1873.¹⁰ Milwaukee, the site of the largest school voucher program in the country, enrolled 17,275 students in the fall of 2006.¹¹ Two new voucher programs were enacted in Arizona in 2006, serving students with disabilities and students in foster care.¹² Georgia also enacted a voucher program for students with disabilities in 2007.¹³

Table 1. Twelve School Voucher Programs in the United States¹⁴

Location	Eligibility	Initiated	Students
Vermont	Rural—no public school	1869	4445
Maine	Rural—no public school	1873	6052
Milwaukee	Means Test	1990	17,275
Cleveland	Means Test	1996	5813
Florida	Disability	1999	16,812
Ohio	Disability, Autism	2003	270
D.C.	Means Test	2004	1718
Utah	Disability	2005	300
Ohio	Needs Improvement Public School	2006	3600
Arizona	Disability	2006	NA
Arizona	Foster Children	2006	NA
Georgia	Disability	2007	NA

* Notes: Enrollments are for the 2006–2007 school year. NA means figures not yet available.

9. FRIEDMAN FOUNDATION, *supra* note 1, at 48; Hammons, *supra* note 8, at 11.

10. FRIEDMAN FOUNDATION, *supra* note 1, at 48; Hammons, *supra* note 8, at 8.

11. FRIEDMAN FOUNDATION, *supra* note 1, at 50. Recently, a new report on the Milwaukee voucher program established that 17,749 voucher students were enrolled in the 122 voucher schools that operated through the entire 2006–2007 school year. See Patrick J. Wolf, THE COMPREHENSIVE LONGITUDINAL EVALUATION OF THE MILWAUKEE PARENTAL CHOICE PROGRAM: SUMMARY OF BASELINE REPORTS 1 (2008), available at http://www.uark.edu/ua/der/SCDF/Milwaukee_Eval/Report_1.pdf. The slightly lower count is presented here because it is drawn from the same source as the counts for other programs that were gathered based on a consistent methodology across voucher programs.

12. FRIEDMAN FOUNDATION, *supra* note 1, at 14, 16.

13. Georgia Department of Education, <http://public.doe.k12.ga.us/sb10.aspx> (follow “Questions & Answers” hyperlink) (last visited Apr. 4, 2008).

14. Information compiled from FRIEDMAN FOUNDATION, *supra* note 1. See also Georgia Department of Education, *supra* note 13.

The incremental trend of establishing additional voucher programs in the 1990s paused from 1999 to 2003 as policymakers awaited the outcome of the constitutional challenge to the Cleveland voucher program. Upon the issuance of *Zelman v. Simmons-Harris*,¹⁵ in which a majority of the Supreme Court upheld the constitutionality of school voucher programs such as Cleveland's, school voucher initiatives re-emerged on the policymaking docket in many states. Whereas only five voucher programs had been established in the 130 years between 1869 and 1999, an additional seven programs were enacted in just the first five years post-*Zelman*.¹⁶

A brief review of the twelve school voucher programs in the United States shows how they were designed to serve exclusively students with various disadvantages. The first two programs in Vermont and Maine were limited to students without public junior high or high schools in their communities.¹⁷ The urban school voucher programs in Milwaukee, Cleveland, and the District of Columbia are restricted to students whose family incomes are at or below 185% of the poverty level.¹⁸ Five statewide voucher programs, including the Carson Smith Scholarship Program in Utah, are limited to students with disabilities.¹⁹ A pioneering voucher program in Arizona is restricted to students in foster care who otherwise would have to change public schools whenever they were placed with a new family.²⁰ Finally, a statewide voucher program in Ohio is limited to students attending schools designated in a state of "academic watch" or "academic emergency."²¹ To even qualify for a school voucher in one of the communities that offer them, a student must have some condition that disadvantages the student vis-à-vis the student's peers.

15. 536 U.S. 639 (2001).

16. See FRIEDMAN FOUNDATION *supra* note 1, at 4; see also *supra* Table 1 (Arizona, District of Columbia, Florida, Georgia, Ohio, and Utah all enacted some kind of voucher program after 2001).

17. Hammons, *supra* note 8, at 5.

18. Specifically, the income ceiling to qualify initially for a voucher is 175% of the poverty level or less in Milwaukee and 185% of the poverty level or less in Cleveland and D.C. FRIEDMAN FOUNDATION, *supra* note 1, at 18, 34, 50.

19. *Id.* at 14, 20, 36, 44 (Arizona, Florida, Ohio, Utah, and Georgia).

20. *Id.* at 16.

21. *Id.* at 38.

The existing school voucher programs deliver on their promise to enroll highly disadvantaged populations of students. For example, over 30% of the students currently served by school vouchers have a diagnosed disability that affects them educationally, which is more than twice the national rate of 14% of K-12 students diagnosed with disabilities.²² John Witte, who led the first official evaluation of a school voucher program in the United States, reported that the Milwaukee Parental Choice Program (MPCP) served disproportionate numbers of students who were low-income, African American, Latino, or who came from single-parent families. Witte wrote, "The MPCP was established and the statute written explicitly to provide an opportunity for relatively poor families to attend private schools. The program clearly accomplished that goal."²³

Like America's first urban school voucher program, the most recent voucher initiative, in the District of Columbia, serves a highly disadvantaged population of students. Over 94% of the students who used a D.C. Opportunity Scholarship (i.e. voucher) during the first year of program operation were African American, compared to 85% of the students in the D.C. public schools (DCPS) who are African American.²⁴ The average family income of initial Opportunity Scholarship Program (OSP) users was \$18,652.²⁵ Eligible applicants to the OSP were significantly more likely to be in special education and also more likely to be enrolled in the federal lunch program for low-income students than non-applicants in DCPS.²⁶ Eligible applicants in the first year of the OSP were performing at achievement levels in reading and math that were statistically similar to non-applicants in the DCPS.²⁷ In a city disproportionately populated by underprivileged children, the D.C. voucher program has attracted and enrolled an especially disadvantaged subgroup of students.

22. U.S. DEP'T OF EDUC., NAT'L CTR. FOR EDUC. STATISTICS, NCES 2006-030, DIGEST OF EDUCATION STATISTICS tbl. 50 (2006).

23. JOHN F. WITTE, *THE MARKET APPROACH TO EDUCATION: EVIDENCE FROM AMERICA'S FIRST VOUCHER PROGRAM*, 58-59 (Princeton University Press 2000).

24. WOLF ET AL., *EVALUATION OF THE DC OPPORTUNITY SCHOLARSHIP PROGRAM: FIRST YEAR REPORT ON PARTICIPATION* (U.S. Department of Education/Institute of Education Sciences 2005) (comparing Table 4-9, fifth row, at. 49, with Table 4-1, third row, at 35).

25. *Id.* at 49.

26. *Id.* at 35.

27. *Id.*

As with the other eleven voucher programs, the D.C. OSP has disproportionately attracted and served highly disadvantaged students *by design*. To be eligible for a voucher, students must live in D.C. and have family incomes below 185% of the poverty level.²⁸ Even the 216 students who were attending private schools when they were awarded vouchers in the first year of the OSP came from highly disadvantaged backgrounds that enabled them to meet the statutory criteria for program eligibility.²⁹ Whenever the program is oversubscribed, which has been the case in all but the first of four years of program operation, public school students attending “needs improvement” schools must be assigned a higher probability of receiving a voucher.³⁰ As Terry Moe discusses in his Article in this volume, statutory instruments of program targeting such as these can be and are regularly used by policymakers to ensure that voucher recipients are less advantaged than the typical K-12 student.³¹ Hard evidence from the twelve voucher programs currently in existence in the United States confirms that targeted vouchers reach students with significant educational needs.

The students who apply for and use vouchers also tend to be educationally disadvantaged because of the logic of parental choice. Commentators can assume, mistakenly, that school choosers must and do engage in “maximizing” behavior that involves an obsessive canvassing of all relevant information and careful consideration of all options. As Herbert Simon observed in his seminal study *Administrative Behavior*, however,

Administrators (and everyone else, for that matter) take into account just a few of the factors of the situation regarded as most relevant and crucial. . . . Because administrators satisfice rather than maximize, . . . they can make their decisions with relatively simple rules of thumb that do not make impossible demands upon their capacity for thought.³²

28. *Id.* at 1.

29. *Id.* at 23. Because the OSP was oversubscribed in the second and all subsequent years of operation, and applicants already attending private school were the lowest programmatic service priority, no additional voucher awards were made to private school applicants after the first year of implementation. *Id.* at 8-9.

30. *Id.* at 19-22.

31. See generally Terry M. Moe, *Beyond the Free Market: The Structure of School Choice*, 2008 BYU L. REV. 557.

32. See HERBERT A. SIMON, *ADMINISTRATIVE BEHAVIOR* 119-20 (The Free Press 4th

When it comes to the education of their children, the simple rule of thumb that parents tend to follow is, "If it ain't broke, don't fix it." Because switching schools is highly disruptive to students—educationally and socially—and requires a significant investment of time and energy by parents, few parents will seek additional schooling options for their child unless they are convinced that the student is underperforming in their current school and that a switch to a different school is likely to generate a significant upside gain. As a result, students perceived by their parents as underperforming will disproportionately comprise the ranks of voucher students.

In summary, school voucher programs are arrangements whereby government funds enable parents to enroll their children in private schools of their choosing. The twelve such programs that currently exist in the United States all are targeted towards student populations that are disadvantaged in one or more ways. Research has confirmed that these voucher programs actually reach their targets. Disadvantaged and underperforming students swell the ranks of voucher programs due to a combination of program design and the logic of parental decision-making. Any universal voucher program that lacks the targeting mechanisms present in all existing voucher programs would be expected to enroll a less disadvantaged population of students.

III. METHODOLOGIES FOR EVALUATING SCHOOL VOUCHERS

The remainder of this Article reviews the evidentiary record surrounding the impacts of school vouchers on the students and parents who seek and use them. Because voucher evaluations present

ed. 2000) (1945), for the classic treatment of the distinction between maximizing and satisficing:

Whereas economic man supposedly maximizes—selects the best alternative from among all those available to him—his cousin, the administrator, satisfices—looks for a course of action that is satisfactory or “good enough” . . . Administrators (and everyone else, for that matter) take into account just a few of the factors of the situation regarded as most relevant and crucial. In particular, they deal with one or a few problems at a time, because the limits on attention simply don't permit everything to be attended to at once. . . . Because administrators satisfice rather than maximize, they can choose without first examining all possible behavior alternatives and without ascertaining that these *are* in fact all the alternatives. Because they treat the world as rather empty and ignore the interrelatedness of all things, . . . they can make their decisions with relatively simple rules of thumb that do not make impossible demands upon their capacity for thought.

significant research challenges, the various methodologies used by voucher researchers are first reviewed and critiqued and only the evidence from the most rigorous and reliable class of studies is presented and discussed.

Even though school voucher programs disproportionately serve highly disadvantaged students, this fact does not allow us to determine the effects of vouchers simply by comparing the outcomes of voucher users to students who do not use vouchers. The parental motivation associated with private school enrollment—with or without a government-financed voucher—could plausibly influence student achievement in the long run independent of the effects of the private school. In methodological terms, simple comparisons of private school students with public school students, voucher applicants with non-applicants, or voucher users with non-users, all will be subject to varying degrees of selection bias. Researchers cannot even be certain of the direction of the bias, as parent motivation to switch their child to a private school may be driven by a sense of desperation, in which case the uncontrolled selection effect will operate in the direction of reporting a negative voucher effect. Parental motivation also may be driven by an inordinate concern for the education of their child, in which case the uncontrolled selection bias will operate in the direction of reporting a positive voucher effect. In either case, the finding would have been false. In the presence of uncontrolled selection bias in education research, analysts simply cannot be confident that any observed difference in the outcomes of program participants relative to non-participants is due to the program and not the selection bias.³³

Education researchers have applied a number of methods in attempts to control for or eliminate selection bias from their evaluations. These approaches can be broadly categorized as cross-sectional studies that statistically model selection, longitudinal studies employing matching techniques, and randomized experiments.

Cross-sectional observational studies that attempt to statistically model selection into private schools provide the weakest protection against selection bias. They are also known as quasi-experimental

33 See generally Robert Boruch, Dorothy De Moya, & Brooke Snyder, *The Importance of Randomized Field Trials in Education and Related Areas*, in EVIDENCE MATTERS: RANDOMIZED TRIALS IN EDUCATION RESEARCH 50-79 (Frederick Mosteller & Robert Boruch eds., 2002).

studies because they use statistical modeling to imperfectly approximate the ideal conditions for identifying causal impacts that actual experiments provide.³⁴ In the case of voucher evaluations, such studies use measurable characteristics of students already attending private schools, at a single point in time, to try to control statistically for the unmeasurable trait of parent motivation and thereby estimate the effects of voucher programs.³⁵ They suffer from three major flaws.

First, estimates of school voucher effects based on cross-sectional analyses use data that are not actually about the question that they seek to answer. Such studies draw evidence from the private sector as a whole, of which over 99% of the students are not voucher users, in order to estimate the likely effects of voucher programs. As discussed above, students who attend private schools using vouchers are, on average, more disadvantaged than the typical public school student and thus dramatically more disadvantaged than the average private school student. Previous research has established that private schooling tends to have larger positive achievement effects on disadvantaged students than on advantaged students.³⁶ Thus, cross-sectional studies of private sector effects on achievement that claim to forecast voucher effects are using data about what is true *in the absence* of voucher students to predict what would be true *for* voucher students. These studies represent the logical equivalent of estimating the effect of a weight-loss program on obese people by studying its effect on normal-weight people. As Henry Levin states, "Of course, none of the public-private comparisons can be as instructive as the direct evaluation of a voucher intervention."³⁷

34. See generally *id.*; WILLIAM R. SHADISH, THOMAS D. COOK & DONALD T. CAMPBELL, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR GENERALIZED CAUSAL INFERENCE* (2001).

35. See generally HENRY BRAUN ET AL., *COMPARING PRIVATE SCHOOLS AND PUBLIC SCHOOLS USING HIERARCHICAL LINEAR MODELING*, NCES 2006-461, available at <http://www.nces.ed.gov/nationsreportcard/pdf/studies/2006461.pdf>; Christopher Lubienski & Sarah Theule Lubienski, *Charter, Private, Public Schools and Academic Achievement: New Evidence from NAEP Mathematics Data* (Nat'l Ctr. for the Study of Privatization in Education, Working Paper Jan. 2006), available at <http://epsl.asu.edu/epru/articles/EPRU-0601-137-OWI.pdf>;

36. See, e.g., JAMES S. COLEMAN & THOMAS HOFFER, *PUBLIC AND PRIVATE HIGH SCHOOLS: THE IMPACT OF COMMUNITIES* (Basic Books 1987).

37. Henry M. Levin, *Educational Vouchers: Effectiveness, Choice, and Costs*, 47 J. POL'Y ANALYSIS & MGMT. 373, 376 (1998).

Moreover, cross-sectional studies of private schooling or school choice typically rely heavily upon participation in federal government aid programs as variables to “control” for selection bias in private-public school comparisons.³⁸ The rates of school-level participation in such programs are much higher in the public sector than in the private sector.³⁹ For example, student disability status is signified in public schools by a student having an Individualized Education Plan (IEP). Students with disabilities who switch to private schools remain disabled but surrender the IEP label.⁴⁰ While the federal government free and reduced price lunch program is offered to students in all public schools, school-level participation in the federal lunch program is discretionary for private schools. Many private schools decline to participate in the federal lunch program because of the extra administrative burden involved.⁴¹ For these reasons, a student with the exact same low income and disability is much more likely to be a participant in the lunch program and have an IEP if he or she attends a public school than if he or she attends a private one.⁴² As a result, modeling selection by including a control variable for participation in the federal lunch program or having an IEP has the practical effect of controlling for the negative effects of low income and disability on test scores among public school students but not among private school students. The predictable effect of such a flawed analytic approach is that the estimate of the “private schooling effect” becomes a negatively biased combination of the true private schooling effect minus the effect of being low income and disabled.⁴³

38. See generally Lubienski & Lubienski, *supra* note 35.

39. GREGORY A. STRIZEK ET AL., NAT'L CTR. FOR EDUC. STATISTICS, CHARACTERISTICS OF SCHOOLS, DISTRICTS, TEACHERS, PRINCIPALS, AND SCHOOL LIBRARIES IN THE UNITED STATES: 2003–2004, at 19–20, 33–34 (2006).

40. See PAUL E. PETERSON & ELENA LLAUDET, PROGRAM ON EDUCATION POLICY AND GOVERNANCE, ON THE PUBLIC-PRIVATE SCHOOL ACHIEVEMENT DEBATE, NCES 2006-313 REVISED 20 (2006) (paper prepared for the annual meetings of the American Political Science Association).

41. *Id.* at 16; see also Lubienski & Lubienski, *supra* note 35, at 22.

42. PETERSON & LLAUDET, *supra* note 40, at 17.

43. For a discussion of the general tendency of analyses of observational data to suffer from negative bias because of measurement limitations, see Robert Boruch, *Encouraging the Flight of Error: Ethical Standards, Evidence Standards, and Randomized Trials*, 113 NEW DIRECTIONS FOR EVALUATION 55, 65–66 (2007).

The third major flaw in cross-sectional analyses of voucher effects is that they are static in that they rely exclusively upon measures of variables at a single point in time. Such studies do not and cannot examine change or growth that is a result of private schooling or school vouchers because their data consists of a single snap-shot of students.⁴⁴ Based on this shortcoming in observational studies, the Charter School Achievement Consensus Panel, a national panel of research experts assembled to evaluate various methods of evaluating school choice interventions such as charter schooling, concluded, “studies using one-year snapshots of achievement cannot have high internal validity, no matter how large a database they draw from or how carefully the analysis is done.”⁴⁵ Robert Boruch sums up the basic weakness of quasi-experimental analyses of cross-sectional data thusly:

Analyses of data from passive surveys or nonrandomized evaluations or quasi-experiments cannot . . . ensure unbiased estimates of the intervention’s relative effect. We cannot ensure unbiased estimates, in the narrow sense of a fair statistical comparison, even when the surveys are conducted well, administrative records are accurate, and analyses of quasi-experimental data are based on thoughtful causal (logic) and econometric models. The risk of misspecified models, including unobserved differences among groups (the omitted variables problem), is often high.⁴⁶

Even perfectly executed analyses of cross-sectional data on private and public schooling are incapable of reliably determining whether or not private schooling has positive, negative, or no effects on students.

Cross-sectional analyses of private schooling effects to estimate voucher effects are the equivalent of using a single photo of a weight-loss client to judge the efficacy of a particular diet compared

44. See PETERSON & LLAUDET, *supra* note 40, at 4–5 (discussing restrictive conditions that must be present in order for a cross-sectional study to produce valid results); Gene V. Glass & Dewayne A. Matthews, *Are Data Enough?*, EDUC. RESEARCHER, Apr. 1991, at 24, 25 (reviewing JOHN E. CHUBB & TERRY M. MOE, POLITICS, MARKETS, AND AMERICA’S SCHOOLS (1990)).

45. JULIAN BETTS & PAUL T. HILL (principal drafters for The Charter School Achievement Consensus Panel), KEY ISSUES IN STUDYING CHARTER SCHOOLS AND ACHIEVEMENT: A REVIEW AND SUGGESTIONS FOR NATIONAL GUIDELINES 3 (Ctr. on Reinventing Public Education, Nat’l Charter Sch. Research Project, NCSRP White Paper Series, Paper No. 2, 2006), available at http://www.ncsrp.org/cs/csr/view/csr_pubs/5.

46. Boruch, *supra* note 43, at 60.

to other diets similarly judged by single photos of clients. The statistical controls included in many of those studies are like air-brushing the photos of the weight-loss clients that you think might have been heavier at the start of their diet—they represent an artificial adjustment based on guesswork. If customers insist on before-and-after photo comparisons at a minimum to evaluate weight-loss programs, shouldn't we insist on at least that much information in evaluating educational interventions such as school vouchers?

Longitudinal studies address the self-selection problem associated with school vouchers by examining changes in student outcomes over multiple time periods. The simplest forms of longitudinal evaluations are individual fixed-effects methods that control for the particular characteristics of study participants by restricting their analysis to variance in the outcomes of the same students over periods when they were and were not exposed to the intervention.⁴⁷ Since the same students are present on both sides of the comparison at different times, student selectivity cannot bias the analysis.⁴⁸ More sophisticated “matching” longitudinal approaches to evaluating school vouchers use information about the characteristics of voucher participants to identify non-participants who “look like voucher participants” in all relevant respects except for voucher participation.⁴⁹ Such voucher-like non-voucher students are described as having a “propensity” to be voucher students, and therefore serve as a more reliable comparison group than just any public school student in evaluations of voucher effects.⁵⁰

Longitudinal studies suffer from none of the major flaws of cross-sectional evaluations of voucher effects.⁵¹ Unlike cross-sectional studies, longitudinal studies examine actual voucher students in estimating what differences, if any, voucher programs make. Longitudinal studies are less subject to bias due to measurement

47. *E.g.*, David N. Figlio & Cecilia Elena Rouse, *Do Accountability and Voucher Threats Improve Low-Performing Schools?* 3–5 (Nat'l Bureau of Econ. Research, Working Paper No. 11597, 2005) (discussing problems with longitudinal studies and formulating a method that addresses these concerns).

48. *See id.* at 5 (noting the benefit of tracking the same students throughout the study).

49. JOHN F. WITTE ET AL., MPCP LONGITUDINAL EDUCATIONAL GROWTH STUDY BASELINE REPORT: SCDP MILWAUKEE EVALUATION REPORT #5, at 8 (2008).

50. *Id.*; Boruch, *supra* note 43, at 64.

51. *See* PETERSON & LLAUDET, *supra* note 40, at 9.

problems since they do not use student characteristics, beyond the identity of students themselves, as explicit controls for selection effects. Finally, longitudinal studies are superior to cross-sectional studies in that they provide evidence of comparative change over time as opposed to mere isolated snapshots of students.⁵² They basically line up the before-and-after photos produced by participants in various weight-loss programs so that the viewer can better evaluate which diet regimen is producing the best results.

Longitudinal studies of school vouchers, though far superior to observational ones, do suffer from one potential flaw. They assume that any effects of selection bias have already influenced the conditions in which students find themselves at the start of the study, but will have little or no influence over the rate of change in those conditions over time. If higher levels of parental motivation really are associated with voucher students, and simultaneously influence how well a student is currently achieving, such unmeasured parental values might presumably influence the rate of change in student achievement as well.⁵³ Thus, longitudinal studies can limit the threat of selection bias to the validity of voucher evaluations, but they cannot eliminate the possibility of such bias entirely.

Random assignment studies can eliminate the threat of selection bias and therefore have justly earned their reputation “as the gold standard for the evaluation of educational interventions” such as voucher programs.⁵⁴ Also known as experiments or randomized controlled trials, random assignment studies take a population of

52. *Id.* (“[S]cholars are most confident of their results when they are able to track student performance over time. Ideally, they prefer four or more observations of the performance of the same student over time, so they can get a sense of the direction a student is moving before and after an educational intervention takes place.”).

53. *See* BETTS & HILL, *supra* note 45, at 12 (“There are a growing number of student-level analyses of trends over time in student test scores that control for individual student characteristics. This represents a far better research design [than observational studies], because it takes into account where a student began on the achievement spectrum and controls for observable student characteristics. However, there remains a risk that a lack of proper controls for unobserved characteristics of each student make (sic) comparisons between students at charters and regular public schools potentially misleading.”) (emphasis omitted).

54. *Id.* at 8; CAROLINE M. HOXBY & SONALI MURARKA, NATIONAL BUREAU OF ECONOMIC RESEARCH, NEW YORK CITY’S CHARTER SCHOOLS OVERALL REPORT, at III.2 (2007); Boruch, De Moya & Snyder, *supra* note 33; *cf.* Thomas D. Cook & Monique R. Payne, *Objecting to the Objections to Using Random Assignment in Educational Research*, in EVIDENCE MATTERS, *supra* note 33, at 150, 174 (affirming the value of the randomized experiment, but acknowledging problems that must be compensated for in the educational context).

equally motivated families and use a random lottery to separate them into a “treatment” group that receives an offer of a voucher and a “control” group that does not receive such an offer.⁵⁵ Because mere chance determined which students are in the treatment and control groups, any differences in educational outcomes subsequently observed between treatment and control students can be reliably attributed to the voucher opportunity as the cause.⁵⁶

Random assignment studies are such powerful instruments for evaluating educational programs that the U.S. Department of Education’s What Works Clearinghouse (WWC) has declared that they are the only research design that meets its evidence standards for rigor “without reservations.”⁵⁷ In contrast, all cross-sectional education studies fail to meet even the minimal WWC evidence standards, and therefore cannot be included in formal reviews of what does and does not work in education, because their lack of both random assignment and baseline data means that we cannot be confident that their treatment and comparison groups were equal in all relevant respects except for the treatment intervention.⁵⁸

Judith Gueron states that a policy experiment “offers unique power in answering the ‘Does it make a difference?’ question. With random assignment, you can know something with much greater certainty and, as a result, can more confidently separate fact from advocacy.”⁵⁹ Robert Boruch describes random assignment evaluations as the modern-day equivalent of the scientific principles of Newtonian physics that Thomas Jefferson described as producing a situation whereby, “Reason and experiment have been indulged, and error has fled before them.”⁶⁰

The high reliability of randomized experiments is the reason why the efficacy of new drugs must be demonstrated in two randomized trials before the Food and Drug Administration permits them to enter the market.⁶¹ Such evaluations are the equivalent of taking a

55. See HOXBY & MURARKA, *supra* note 54, at III.2.

56. Robert Boruch, *The Null Hypothesis Is Not Called That for Nothing: Statistical Tests in Randomized Trials*, 3 J. EXPERIMENTAL CRIMINOLOGY 1, 2; Boruch, *supra* note 43, at 60.

57. WHAT WORKS CLEARINGHOUSE, EVIDENCE STANDARDS FOR REVIEWING STUDIES 1, 5, 11 (2006), available at http://ies.ed.gov/ncee/wwc/pdf/study_standards_final.pdf.

58. *Id.*

59. Judith M. Gueron, *The Politics of Random Assignment: Implementing Studies and Affecting Policy*, in EVIDENCE MATTERS, *supra* note 33, at 15.

60. Boruch, *supra* note 43, at 55.

61. Boruch, *supra* note 56, at 11.

group of over-weight people who all want to lose weight, using a lottery to determine which ones will receive a supervised administration of the Nutrisystem diet and which ones will be left to their own devices, and calculating the average weight-loss for the two groups at a later point in time. Because random assignment approximately equalizes groups on both measurable and unmeasurable characteristics, we could confidently attribute any significantly higher or lower level of weight-loss among the treatment group to the Nutrisystem intervention.

Even though random assignment studies are widely revered in medicine, economics, political science, and education, some researchers have recently raised claims that such experimental studies suffer various biases.⁶² These researchers claim that since some members of the treatment group inevitably decline to use vouchers and some members of the control group obtain private schooling without the assistance of a voucher, the estimates of the program suffer from “compliance/attrition bias.”⁶³ This claim stems from a basic misunderstanding of the logic of experimental program evaluations. Public policies cannot force clients to use programs. They can only offer the services to qualified clients. A public policy can fail to produce significant outcomes, either because it is not effective if used, or because it is effective but low percentages of clients use it consistently. As such, the outcomes that a policy intervention like vouchers generates for non-users should be and typically are averaged into the effects that it produces for voucher users, thereby producing an accurate and unbiased estimate of the impact of the offer of a voucher, which is all that public policy can provide. Similarly, control group members that obtain the equivalent of the voucher “treatment” without the assistance of the voucher offer are an authentic part of the control group counterfactual. We know that, absent a voucher program, those students would have attended private school anyway, because their actual behavior has revealed this to us. Control group “crossover” to a treatment-like condition is thus not a source of bias in experimental analyses.

62. See Christopher Lubienski & Peter Weitzel, *The Effects of Vouchers and Private Schools in Improving Academic Achievement: A Critique of Advocacy Research*, 2008 BYU L. REV. 447, 475-78.

63. *Id.* at 477.

The fact that some students randomly offered vouchers do not use them, and some control group members attend private schools without vouchers, does not in any way bias the estimate of the impact of offering students vouchers, though it does generate a conservative estimate of the effects of actually attending private school.⁶⁴ This is because the outcomes for voucher decliners, for whom the impact of the voucher is zero, are averaged in with those of voucher users in calculating the experimental impact of vouchers. Likewise, any change in outcomes experienced by control group members who attend private schools are included in an experimental analysis on the control-group side of the comparison. If analysts or policymakers want to draw from an experimental evaluation in determining the impact of actual voucher usage or private schooling, established statistical techniques exist and are regularly employed in experimental voucher evaluations to produce unbiased estimates of those impacts.⁶⁵ So, if one is interested in the average effects of a program that merely offers students vouchers, random assignment studies, as traditionally implemented, generate unbiased estimates of that average “intend-to-treat” impact. If one is instead interested in the average effect of obtaining the actual experience that vouchers are supposed to enable students to receive—private schooling—then established statistical methods exist that can be and are applied to experimental voucher data to produce unbiased estimates of private schooling either through voucher usage or in general. The most commonly used such method is Instrumental Variable (IV) analysis with the original voucher lottery as the ideal instrument.⁶⁶

Second, some researchers claim that experimental evaluations of voucher impacts suffer from “generalizability” bias because the populations of students who choose to apply for voucher programs are different from non-applicants. It is true that the results of any particular experimental voucher evaluation only strictly apply to the special conditions in which the program was designed and implemented. It would be risky to claim that the results of an experimental voucher evaluation of a means-tested inner-city

64. See WILLIAM G. HOWELL & PAUL E. PETERSON, *THE EDUCATION GAP* 49 (2002).

65. *Id.* at 49–51; PATRICK WOLF ET AL., INST. OF EDUC. SCIENCES, *EVALUATION OF THE DC OPPORTUNITY SCHOLARSHIP PROGRAM: IMPACTS AFTER ONE YEAR* 67–79 (2007).

66. HOWELL & PETERSON, *supra* note 64, at 50; Joshua D. Angrist, Guido W. Imbens & Donald B. Rubin, *Identification of Causal Effects Using Instrumental Variables*, 91 J. AM. STAT. ASS'N. 444, 444–45 (1996).

program would automatically apply to a statewide voucher program for students of any income level but with disabilities. Those are two populations that differ in ways that could plausibly influence their response to vouchers, so analysts should not, and generally do not, make such generalizability claims. This condition is not, properly understood, a “bias” of experimental voucher studies, since it does not undermine the validity of experimental impact estimates. It is simply a limitation.

Experimental evaluations are purposely designed to be exceptionally strong in their “internal validity”—that is in their ability to reach an accurate determination as to whether or not the voucher program impacted a certain group of study participants.⁶⁷ Experiments of all types are inherently limited in their “external validity,” that is, in the ability to apply the results of one study of a particular student population to the context of a very different student population. Presumably, one need determine with confidence whether or not an educational intervention works with a given set of students before one considers whether it might work for a different group of, or all types of, students. That is why experimental evaluators of voucher programs qualify their findings to make it clear that different results could emerge from similar evaluations of very different student populations.⁶⁸ That is not a bias, just good scholarly practice.

In a broader sense, the charge that voucher evaluations are biased because they are limited to families interested in applying for vouchers borders on the ridiculous. True, such evaluations only tell us what impact the program will have on families who want to use it. What else would policymakers want to know? Certainly it would be of little value to know what impact voucher programs have on families who do not want to and never will use them. Would we care what effect a particular diet program had on people who do not want to lose weight?

Finally, some researchers claim that experimental evaluations of voucher programs are biased by the fact that the peer groups in private schools are different from those in public schools.⁶⁹ If, in fact,

67. See Cook & Payne, *supra* note 54, at 150, 150–57, 164.

68. HOWELL & PETERSON, *supra* note 64, at 204–05 (emphasizing assumptions upon which conclusions are based).

69. Lubienski & Lubienski, *supra* note 35, at 40; see also PETERSON & LLAUDET, *supra* note 40, at 7.

the backgrounds of students in voucher-participating private schools are more advantaged on average than those of students in the schools attended by control-group students, then that is a legitimate aspect of the treatment. Voucher programs enable students to switch from one type of educational environment to another one. If that new educational environment, in a participating private school, is different, for example because it includes more high-income peers, then that is part and parcel of the treatment. Analysts must not control for the differing characteristics of peer groups in the schools being attended by students in experimental voucher evaluations because doing so “controls away” one of the legitimate sources of any treatment effect. If voucher students learn more because they are surrounded by more advantaged peers in their new schools, then that is an explanation for why vouchers work, not something that should be subtracted out from any calculation of whether or not vouchers work.

Our weight-loss example is again instructive. When comparing the effectiveness of different weight-loss programs, should we control for the caloric intake of their prescribed menus? Researchers who argue that experimental voucher studies need to control for peer-group effects would have to similarly claim that weight-loss program comparisons would have to control for the fitness of the people in the weight-loss candidate’s support group. Their claim would be that a given program is not more effective—it merely surrounds the over-weight person with more fit and inspiring peers. The legitimate defense of such a weight-loss program is that it may be more effective than alternative programs *because* fit support group members inspire participants to lose weight, but the specific reason for its greater effectiveness does not change the simple fact that it *is* more effective. The relative fitness of other weight-loss participants is part of the treatment package, not a biasing factor, just as differing peer-group characteristics are a legitimate part of the voucher treatment and not something that should be netted out of the equation.

Because experimental voucher evaluations are rightly viewed as the gold standard of evaluation, and the claims of bias raised against them do not survive close scrutiny, here we confine our examination of the impacts of vouchers to the growing body of ten analyses that

meet this highest of standards for rigor.⁷⁰ All ten of these studies appeared in reports that survived peer review prior to their public release or publication.⁷¹ They paint a modestly positive picture of the impacts of school vouchers on parent and student outcomes.

IV. VOUCHER IMPACTS AS REPORTED IN RANDOM ASSIGNMENT STUDIES

The random assignment studies of actual school voucher programs in the United States indicate that they have consistently large positive effects on parental satisfaction with schools and smaller and less consistent effects—but always positive—on student test scores. Voucher programs demonstrate their most immediate and largest positive impacts on the expressed levels of satisfaction that parents have with their child's school. This school voucher impact has been confirmed in all five random assignment studies that

70. See *infra* Table 2.

71. Eight of the ten random assignment studies of voucher impacts discussed here were published in conventional peer-review academic journals. Several of the publication venues were the most prestigious scientific journals in their respective fields. Those journals were *The American Behavioral Scientist*, *Education and Urban Society*, *Journal of the American Statistical Association*, *Journal of Policy Analysis and Management*, *Policy Studies Journal*, and *Quarterly Journal of Economics*. John Barnard et al., *Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City*, 98 J. AM. STAT. ASS'N 299 (2003); Joshua M. Cowen, *School Choice as a Latent Variable: Estimating the 'Complier Average Causal Effect' of Vouchers in Charlotte*, 36 POLY STUD. J. (forthcoming 2008); Jay P. Greene et al., *Effectiveness of School Choice: The Milwaukee Experiment*, 31 EDUC. & URB. SOC'Y 190 (1999) [hereinafter Greene et al. (1999)]; William G. Howell et al., *School Vouchers and Academic Performance: Results from Three Randomized Field Trials*, 21 J. POLY ANALYSIS & MGMT (2002); Alan B. Krueger & Pei Zhu, *Another Look at the New York City School Voucher Experiment*, 47 AM. BEHAV. SCIENTIST 658 (2004) [hereinafter Krueger & Zhu (2004a)]; Alan B. Krueger & Pei Zhu, *Inefficiency, Subsample Selection Bias, and Nonrobustness: A Response to Paul E. Peterson and William G. Howell*, 47 AM. BEHAV. SCIENTIST 718 (2004) [hereinafter Krueger & Zhu (2004b)]; Paul E. Peterson & William G. Howell, *Efficiency, Bias, and Classification Schemes: A Response to Alan B. Krueger and Pei Zhu*, 47 AM. BEHAV. SCIENTIST 699 (2004); Cecilia Elena Rouse, *Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program*, 1998 Q.J. ECON. 553. One study was published in the research section of the journal *Education Matters* (now called *Education Next*). Jay P. Greene, *Vouchers in Charlotte*, 1 EDUC. MATTERS 55–60 (2001) [hereinafter Greene (2001)]. Although some articles in *Education Matters/Education Next* are commentaries that are not subject to peer review, all articles published in the research section of the journal must survive peer-review, so that study was in fact peer-reviewed prior to publication. The final random assignment study discussed here was subject to a thorough external peer review before being released as an official federal government report. WOLF ET AL., *supra* note 65.

explored the question of parental satisfaction.⁷² Voucher programs appear especially to increase parent satisfaction regarding curriculum, safety, parent-teacher relations, academics, and the religious environment of schools.⁷³ The positive impacts of voucher programs on parental satisfaction are large, averaging three-tenths of a standard deviation, or more than one-third of the size of the notorious black-white test score gap.⁷⁴ As an example, seventy-four percent of parents of students offered a voucher in the new District of Columbia Opportunity Scholarship Program graded their child's school "A" or "B," compared to just fifty-five percent of the control group.⁷⁵

Although it is indisputable that parents are more satisfied with their child's school if they have been given a voucher, we do not yet know why they are so much more satisfied. The private schools that parents select using vouchers might be more effective schools that do a better job educating students. Voucher parents might be more satisfied with schools because they are a more comfortable environment for their child, in terms of safety and programs, than their previous public school—even if the voucher schools do no better than public schools at educating students. Finally, the large impacts that voucher programs have on subsequent parental satisfaction might be the result of cognitive dissonance. Since parents themselves selected their child's new school, they might feel vested in the outcome of the choice and filter their perceptions in such ways that the voucher schools look better to them even if, objectively, they are no better than the child's previous public school. Voucher studies have yet to determine if greater parental satisfaction with voucher-accepting private schools is grounded in fact or false perception, but central to that consideration is the impact that vouchers have on student achievement.

A substantial base of evidence is emerging that indicates that school voucher programs tend to boost the achievement of some or all of the students who use them. As discussed *supra*, enough gold-standard random-assignment studies have been completed that we

72. WOLF ET AL., *supra* note 65, at xiii-xxii; HOWELL & PETERSON, *supra* note 64, at 39-52; Greene (2001), *supra* note 71.

73. WOLF ET AL., *supra* note 65, at I.3; HOWELL & PETERSON, *supra* note 64, at 179-84.

74. See WOLF ET AL., *supra* note 65, at I.3; HOWELL & PETERSON, *supra* note 64, at 179-84.

75. See WOLF ET AL., *supra* note 65, at 56-57.

can and should limit our consideration exclusively to those highly rigorous evaluations. A total of eight different research teams have conducted ten separate analyses of data produced by six random assignment voucher programs in five different cities.⁷⁶ Three of the analyses—two in Milwaukee and one in the District of Columbia—had full-tuition, publicly funded school vouchers as the subject of the evaluation.⁷⁷ The other seven analyses were of data from voucher-like, privately funded, partial-tuition scholarship programs in Charlotte, Dayton, D.C., and New York City.⁷⁸ All ten of the analyses examined the impacts of using a voucher on subsequent student achievement measured by test scores in reading and math, either separately or combined into a single composite score.⁷⁹ All of the studies, except for the re-analysis of the New York City experiment by Krueger and Zhu, used the combination of random assignment data and well-established statistical methods to produce unbiased estimates of the educational impacts of actually using a voucher or voucher-like scholarship. What follows is a brief review of the studies:

76. See *infra* Table 2.

77. WOLF ET AL., *supra* note 65, at xiii-xvi (noting, however, that 8.4% of the students in the study group “were attending a school charging tuition above the statutory cap of \$7500 in their first year in the Program”); Greene et al. (1999), *supra* note 71, at 190, 194; see Rouse, *supra* note 71, at 554.

78. Barnard et al., *supra* note 71, at 299; Cowen, *supra* note 71; Greene (2001), *supra* note 71, at 55; Howell et al., *supra* note 71, at 194; Krueger & Zhu (2004a), *supra* note 71, at 658-59; Krueger & Zhu (2004b), *supra* note 71, at 718; Peterson & Howell, *supra* note 71, at 699.

79. WOLF ET AL., *supra* note 65, at 44-45; Barnard et al., *supra* note 71, at 306; Cowen, *supra* note 71; Greene (2001), *supra* note 71, at 57; Greene et al. (1999), *supra* note 71, at 200-01; Howell et al., *supra* note 71, at 200; Krueger & Zhu (2004a), *supra* note 71, at 659-60; Krueger & Zhu (2004b), *supra* note 71, at 726-27; Peterson & Howell, *supra* note 71, at 699; Rouse, *supra* note 71, at 575-77.

Table 2. Summary of Impacts of Voucher Usage from Random Assignment Studies

Location	Author (Year)	Math Impacts	Reading Impacts	Subgroup Impacts	Notes/Data Span/Grades
Charlotte, NC	Greene (2001)	5.9 NPR Average Gain	6.5 NPR Average Gain	NA	Voucher only a portion of tuition. 1 year, Grades 2-8
Charlotte, NC	Cowen (Forthcoming)	4-6 NPR Average Gain	5-8 NPR Average Gain	NA	Estimates sensitive to models of treatment compliance; 5 of 6 significant at $p < 0.10$
Dayton, OH	Howell et al. (2002)	No significant gains (subjects combined)		6.5 NPR gain for African Americans, 2d year	Voucher only a portion of tuition. 2 years, Grades 1-7
District of Columbia	Howell et al. (2002)	7.5 NPR gains only in 2 nd year (subjects combined)		9.2 NPR gains for African Americans 2d year	Voucher only a portion of tuition. 3 years, Grades 1-8
District of Columbia	Wolf et al. (2007)	No significant gains	No significant gains	7.8 SS gain for non-SINI students in math, 1st year 6.7 SS gain for higher baseline performers in math, 1st year	Full voucher, Grades K-12

Milwaukee, WI	Greene et. al. (1999)	10.7 NPR gain after 4 years, (5 NPR after 3 years)	5.8 NPR gain after 4 years, (2-3 NPR after 3 years)	NA	Test scores grow most in the later years. Secular schools only
Milwaukee, WI	Rouse (1998)	1.5-2.3 NPR Gain <i>per year</i>	No significant gains	NA	Secular schools only
New York, NY	Barnard et. al. (2003)	No significant gains	No significant gains	4-6 NPR gains in math for African Americans and those from lower-performing schools	1 year. Grades 1-4
New York, NY	Howell et al. (2002)	No significant gains		4.3-9.2 NPR gains for African-Americans	3 years. Grades 1-4
New York, NY	Krueger and Zhu (2004)	No significant gains		No significant gains	3 years. Grades K-4

* Notes: NA means "not attempted;" NPR means National Percentile Rank; SS means Scale Score; and non-SINI means student did not attend a "school in need of improvement" under NCLB at the time of application.

Two independent studies have been produced using experimental data from the privately funded, partial-tuition, voucher-like scholarship program in Charlotte, North Carolina. Jay P. Greene of the University of Arkansas conducted the original Charlotte study.⁸⁰ Of 1143 eligible students who had entered a scholarship lottery and had been randomly assigned to the treatment or control groups, a total of 452 (40%) in grades two through eight produced outcome data one year after random assignment.⁸¹ Using Instrumental Variable (IV) analysis to correct for the differences

80. Greene (2001), *supra* note 71.

81. *Id.* at 56.

between users and non-users among the treatment group, Greene reports statistically significant achievement gains of 5.9 percentiles in math and 6.5 percentiles in reading for the voucher students compared to the control group after one year.⁸² This study was published in the peer-reviewed section of the journal *Education Matters*, which later became *Education Next*, rated by *Education Week* as the most influential education policy journal in the United States.

Joshua Cowen of the University of Wisconsin has re-analyzed the data from the Charlotte voucher experiment and largely replicates Greene's original results.⁸³ Cowen uses a variety of new maximum likelihood statistical techniques in place of IV analysis to generate unbiased estimates of achievement effects of the voucher program on voucher users, which he calls the "complier average causal effect."⁸⁴ Cowen reports voucher achievement gains of 4–6 percentiles in math and 5–8 percentiles in reading that are statistically significant with at least 90% confidence in five of the six alternative regression models that he estimates.⁸⁵

William Howell of the University of Chicago led a research team that evaluated the impact of privately funded, partial-tuition, voucher-like scholarship programs in Dayton, the District of Columbia, and New York City.⁸⁶ In Dayton, Ohio, 56% of the 803 randomized voucher applicants turned out for outcome data collection one year after and 49% turned out two years after random assignment.⁸⁷ The authors report the simple "intent-to-treat" results of being offered a voucher, regardless of whether or not it is used, and use IV analysis to generate unbiased estimates of the effects of actual scholarship usage.⁸⁸ Howell and his colleagues report no statistically significant achievement gains for the group treated with vouchers in Dayton in years one or two.⁸⁹ They do, however, report statistically significant achievement gains from voucher usage for the

82. *Id.* at 57.

83. Cowen, *supra* note 71.

84. *Id.*

85. *Id.*

86. Howell et al., *supra* note 71, at 191–217; William G. Howell et al., *Vouchers in New York, Dayton, and D.C.*, 2 *EDUC. MATTERS* 48–54 (2001).

87. Howell et al., *supra* note 71, at 195.

88. *See id.* at 194–98, 202–06.

89. *Id.* at 200.

African American subgroup of students in their Dayton sample of 6.5 percentiles in math and reading combined in the second year.⁹⁰ The results of the Dayton voucher experiment were published in the *Journal of Policy Analysis and Management*, which is the leading peer-reviewed public policy journal in the United States, as well as the peer-reviewed research section of *Education Matters*.

Howell and his research team also reported the results of a similar analysis of the impact of privately funded, partial-tuition, voucher-like scholarships in Washington, D.C. A total of 1582 students were randomly assigned to treatment and control in the initial D.C. voucher experiment.⁹¹ Sixty-three percent of participants turned out for outcome data collection in year one and 50% in year two.⁹² The research team reported no statistically significant general impacts of voucher usage in year one but observed significant gains of 7.5 percentiles in combined math and reading achievement for the D.C. students treated with vouchers by year two.⁹³ In the subgroup of students in the D.C. study who were African American—over 90% of the study sample—the voucher gains in the second year totaled 9.2 percentiles.⁹⁴ These results of the initial D.C. voucher experiment appeared in the same two peer-reviewed journals as the Dayton results discussed above as well as a book by the Brookings Institution Press in its second edition.⁹⁵

The initial D.C. voucher experiment was not the final word on voucher impacts in the nation's capital. In January of 2004, Congress passed and President Bush signed into law the first federally funded school voucher program.⁹⁶ The Opportunity

90. *Id.* at 205.

91. *Id.* at 195.

92. *Id.*

93. WILLIAM G. HOWELL AND PAUL E. PETERSON WITH PATRICK J. WOLF AND DAVID E. CAMPBELL, *THE EDUCATION GAP: VOUCHERS AND URBAN SCHOOLS* 146 (Brookings Institution Press rev. ed. 2006). The voucher impacts observed in Washington, D.C., from the same study reported in HOWELL & PETERSON, *supra* note 64, at 205, are slightly larger because they are limited to participants who were African American.

94. HOWELL ET AL., *supra* 93, at 205.

95. Howell et al., *supra* note 71, at 205. The D.C. experiment was continued for a third outcome year, during which large numbers of both the treatment and control groups transferred to D.C. charter schools. With high percentages of both the treatment and control group students attending the same charter schools that year, the experiment essentially collapsed and no significant voucher impacts on test scores were observed. HOWELL & PETERSON, *supra* note 64, at 146–50.

96. See Rod Paige & Anthony Williams, *Working Together To Offer a Choice*, WASH.

Scholarship Program is being implemented in Washington, D.C., and evaluated using rigorous experimental methods.⁹⁷ A total of 2308 low-income D.C. students in two cohorts (2004 and 2005) were randomly assigned to receive a voucher worth up to \$7500 or serve in the control group.⁹⁸ Effectively 77% of this large sample of study participants turned out for outcome data collection one year after random assignment.⁹⁹ The evaluation team reported no overall test score gains due to the vouchers in the first outcome year.¹⁰⁰ They did report achievement gains in math of 7.8 scale score points for voucher students who previously had been attending lower-performing public schools as well as math gains of 6.7 scale score points for voucher students whose baseline test score performance was in the upper two-thirds of the overall low test score distribution of the students in the sample.¹⁰¹ This evaluation, supervised by the U.S. Department of Education's Institute for Education Sciences, is ongoing.

Two independent research teams produced separate analyses of experimental data from the initial evaluation of the Milwaukee Parental Choice (voucher) Program (MPCP), originally conducted by John Witte of the University of Wisconsin. Witte's evaluation used longitudinal and not experimental methods.¹⁰² Voucher lotteries were used in the MPCP in its early years, however, allowing subsequent researchers to employ experimental methods to analyze the data.¹⁰³ The first experimental study of the Milwaukee program, by Jay Greene and his colleagues, reported statistically significant voucher impacts on both math and reading test scores that were modest for three years after random assignment but moderately large after four years.¹⁰⁴ The researchers reported no significant voucher impacts on test scores until students had used them for at least three

TIMES, Feb. 5, 2004, at A21.

97. See generally WOLF ET AL., *supra* note 65; WOLF ET AL., *supra* note 24.

98. WOLF ET AL., *supra* note 65, at xiii-xiv.

99. *Id.* at F-5.

100. *Id.* at 44-45.

101. See *id.* at 74.

102. JOHN F. WITTE, THE MARKET APPROACH TO EDUCATION: EVIDENCE FROM AMERICA'S FIRST VOUCHER PROGRAM 122-32 (2000).

103. Greene (2001), *supra* note 71, at 55.

104. See generally Greene et al. (1999), *supra* note 71.

years.¹⁰⁵ Their study was published in the peer-reviewed journal *Education and Urban Society*.

In a separate analysis, Cecilia Rouse of Princeton University largely replicated the Greene et al. Milwaukee study.¹⁰⁶ She used “years of voucher use” as her explanatory variable—instead of looking at impacts in separate years—concluding that the Milwaukee voucher program generated math gains of 1.5 to 2.3 percentiles per year but no statistically significant reading gains.¹⁰⁷ Rouse’s replication study was published in the peer-reviewed *Quarterly Journal of Economics*.

Finally, three different research teams have analyzed random assignment data from the New York City experiment with privately funded, partial-tuition, voucher-like scholarships.¹⁰⁸ All three studies reported no statistically significant achievement gains overall due to the vouchers. Two of the three studies, however, reported voucher achievement gains for one or more subgroups of study participants.

John Barnard, a research statistician at deCODE Genetics, and his colleagues produced the most optimistic assessment of the test score impacts of the New York City voucher experiment. Using propensity-score matching techniques instead of IV analysis to generate unbiased estimates of the impact of voucher usage on student achievement one year after random assignment, Barnard and his colleagues found no statistically significant gains overall.¹⁰⁹ They did, however, report statistically significant voucher gains in math of 4–6 percentiles for African Americans and students who previously were attending lower-performing schools.¹¹⁰ Their study was published in the peer-reviewed *Journal of the American Statistical Association*, the top-ranked statistics journal in the United States.

Howell and his colleagues published the first reports drawing from data collected on the New York City voucher experiment.¹¹¹ Their results are almost identical to those reported by Barnard et al. Using IV analysis to generate unbiased estimates of the impacts of

105. *Id.* at 200.

106. *See generally* Rouse, *supra* note 71.

107. *See id.* at 576–77, 593.

108. *See generally* Barnard et al., *supra* note 71; Howell et al., *supra* note 71; Krueger & Zhu (2004b), *supra* note 71.

109. *See* Barnard, *supra* note 71, at 308–09.

110. *See id.*

111. *See generally* Howell et al., *supra* note 71.

voucher use, Howell et al. found no statistically significant test score impacts overall but did report statistically significant achievement gains for African Americans.¹¹² The voucher gains for African Americans were significant in all three years of the evaluation and ranged from 4.3 to 9.2 percentiles.¹¹³ The results from the first two years of this study were published in the peer-reviewed *Journal of Policy Analysis and Management* and the peer-reviewed research section of *Education Matters*. The results from all three years were published in a book, now in its second edition, by the Brookings Institution Press.¹¹⁴

Finally, Alan Krueger and Pei Zhu of Princeton University conducted the third analysis of the New York City experimental data.¹¹⁵ By adding test scores from kindergartners who were not tested at baseline, reclassifying the races of study participants, and including a more extensive set of baseline variables in their statistical models, Krueger and Zhu generate estimations of voucher impacts that are still positive but are not statistically significant overall or specifically for African Americans.¹¹⁶ Their unconventional approach to analyzing these data has been the subject of heated controversy.¹¹⁷ Their study was published in the peer-reviewed journal *The American Behavioral Scientist*.

None of the experimental analyses of voucher effects on student achievement reports exactly the same results. Nevertheless, a careful examination of the evidentiary record to date reveals some general patterns of outcomes. Nine of the ten gold standard evaluations of voucher programs have reported positive and statistically significant achievement impacts for all or at least some subgroup of voucher recipients.¹¹⁸ Five of the ten analyses concluded that all types of

112. *Id.* at 212.

113. HOWELL & PETERSON, *supra* note 63, at 146.

114. *Id.* at 144–50.

115. See generally Krueger & Zhu (2004b), *supra* note 71.

116. *Id.* at 693–95.

117. See generally William G. Howell & Paul E. Peterson, *The Use of Theory in Randomized Field Trials: Lessons from School Voucher Research on Disaggregation, Missing Data, and the Generalization of Findings*, 47 AM. BEHAV. SCIENTIST 634 (2004); Krueger & Zhu (2004a), *supra* note 71; Krueger & Zhu (2004b), *supra* note 71; Peterson & Howell, *supra* note 71.

118. WOLF ET AL., *supra* note 65, at 44–49; Barnard et al., *supra* note 71, at 299, 308–09; Cowen, *supra* note 71; Greene (2001), *supra* note 71, at 57; Greene et al. (1999), *supra* note 71, at 200; Howell et al., *supra* note 71, at 191; Krueger & Zhu (2004a), *supra* note 71,

participants benefited academically from a school voucher.¹¹⁹ Of the five other studies that did not report a significant “general” voucher effect on test scores, four of them reported clear voucher achievement gains for at least one major subgroup of participants.¹²⁰ Only one of the ten studies—the re-analysis by Krueger and Zhu of the earlier New York City experiment—concluded that a voucher program had no clear achievement benefits for any group of participants.¹²¹ No random assignment study of vouchers to date has indicated that vouchers harm students academically.

The results of random assignment studies of school vouchers reveal more than simply a general tendency for vouchers to boost student achievement. The pattern of experimental results suggests that achievement gains from using a voucher are more common in math than in reading. This finding is not surprising, given that math achievement is more heavily a function of school instruction than is reading achievement. Educational achievement gains from vouchers appear to be largest and most consistent for African American students, the ethnic category of students long recognized as being most disadvantaged by residential assignment to poorly performing public schools.¹²² Voucher-induced test score gains for all or some of the study participants were apparent in the first outcome year of experimental studies in Charlotte, New York City, and the second experiment in the District of Columbia.¹²³ For the voucher experiments in Dayton, the original D.C. program, and Milwaukee, positive and statistically significant voucher impacts on student test scores did not emerge until two or more years after students switched schools using the voucher.¹²⁴ Since school switching is a

at 693; Peterson & Howell, *supra* note 71, at 699, 715; Rouse, *supra* note 71, at 575–78.

119. Greene (2001), *supra* note 71, at 57; Greene et al. (1999), *supra* note 71, at 200; Howell et al., *supra* note 71, at 205 (but only for D.C.) (repeated in Howell et al., *Vouchers in New York*, *supra* note 86, at 146 (again only for D.C.)); Rouse, *supra* note 71, at 575–78.

120. WOLF ET AL., *supra* note 65, at 44–49; Barnard et al., *supra* note 71, at 299, 308–09; Howell et al., *supra* note 71, at 191; Peterson & Howell, *supra* note 71, at 699, 715.

121. Krueger & Zhu (2004b), *supra* note 71, at 726–27.

122. See, e.g., Derek Neal, *The Effects of Catholic Secondary Schooling on Educational Achievement*, 15 J. LAB. ECON. 100 (1997).

123. Barnard et al., *supra* note 71, at 299, 308–09; Cowen, *supra* note 71; Greene (2001), *supra* note 71, at 57; Howell et al., *supra* note 71, at 200; Krueger & Zhu (2004a), *supra* note 71, at 693; Peterson & Howell, *supra* note 71, at 699.

124. WOLF ET AL., *supra* note 65, at 44–49; Greene et al. (1999), *supra* note 71, at 200; Rouse, *supra* note 71, at 575–78.

necessary element of voucher use known to temporarily disrupt student learning,¹²⁵ it is not surprising that voucher test score gains were somewhat slow to appear in several of the experimental studies.

Researchers have been studying the effects of school voucher programs on participating students and parents for more than a decade. A total of ten gold-standard, peer-reviewed experimental studies have been produced thus far, demonstrating conclusively that school vouchers increase parental satisfaction with schools and providing substantial evidence that at least some students are helped academically by vouchers.¹²⁶ More high quality experimental research is needed before we can close the books on the participant effects of school vouchers, but the results to date are generally promising.

It is important to acknowledge that all previous studies of school vouchers and voucher-like private scholarships involved programs targeted to low-income, inner-city students. Public opinion surveys regularly indicate that urban, minority, and low-income parents are more supportive of school vouchers than the general public,¹²⁷ and policy makers have responded by targeting voucher programs to serve the constituency that is most loudly calling for them.¹²⁸ The research findings reviewed here are characteristic of the targeted voucher programs that currently exist in the United States. We cannot be certain that similar outcomes would be generated by a universal voucher program such as the one that failed the recent referendum vote in Utah.

125. Erik A. Hanushek et al., *Disruption Versus Tiebout Improvement: The Costs and Benefits of Switching Schools*, 88 J. PUB. ECON. 1744 (2004); Judy A. Temple & Arthur J. Reynolds, *School Mobility and Achievement: Longitudinal Findings from an Urban Cohort*, 37 J. SCH. PSYCHOL. 355-77 (1999).

126. See *supra* notes 70, 71-79 and accompanying text and table.

127. INSTITUTE FOR JUSTICE, PUBLIC OPINION DATA: AMERICANS WANT SCHOOL CHOICE 3-4 (2004), available at http://www.ij.org/pdf_folder/school_choice/public_opinion_data.pdf; David L. Leal, *Latinos and School Vouchers: Testing the "Minority Support" Hypothesis*, 85 SOC. SCI. Q. 1227, 1236 (2004); *Support for School Choice Spans Political Spectrum*, 1 ADVISOR (2002), available at http://schoolchoicewi.org/news/index.cfm?nt_id=1&cn_id=12#anchor0 ("Nationally, the strongest support comes from low-income parents, especially African Americans and Hispanics.").

128. See, e.g., DEBRA HARE FITZPATRICK & ILENE BERMAN, CENTER FOR SCHOOL CHANGE, PROVIDING QUALITY CHOICE OPTIONS IN EDUCATION 9 (2005), available at http://www.allianceforschoolchoice.org/research_school_choice.aspx?IIID=24&IIID=2339.

V. CONCLUSION

Much is already known with confidence regarding school voucher programs in the United States. The most reliable information about such programs has been generated by way of ten gold-standard analyses of random assignment voucher or voucher-like experiments. We know that they are targeted to underprivileged children and, as a result, disproportionately serve students that are highly disadvantaged. We know that parents are much more satisfied with their child's school if they have used a voucher to choose it. We know, through the assistance of a substantial body of rigorous experimental studies, that the effect of vouchers on student achievement tends to be positive; however, achievement impacts are not statistically significant for all students in all studies and they tend to require several years to materialize. The existing research base, however, tells us nothing with certainty about what would happen were school vouchers offered to all students of a particular state or our nation. Such a policy proposal would be a voyage into the unknown. The voucher journey targeted to disadvantaged students is well-charted and promising. The evidence to date suggests that policymakers are relatively safe in traveling that course.

Policy makers are also urged to increase support for randomized trials to evaluate controversial education interventions such as school vouchers. Eschewing randomized trials in education research, as Boruch observes, leaves "the great questions of society to the ignorant advocates of change on the one hand and ignorant opponents of change on the other."¹²⁹ For the sake of the next generation of children, we can and should do better.

129. Boruch, *supra* note 56, at 19.