The third-year evaluation of the federally funded Washington, D.C. voucher program shows that low-income students offered vouchers in the first two years of the program had modestly higher reading scores after three years but showed no significant difference in mathematics. Students were randomly assigned to treatment and control groups, and the authors assessed the treatment effect on the overall, combined sample as well as some sample subgroups. The authors, however, interpret the results in ways that raise questions given some of their own findings. For instance, the report downplays the implications of the subgroup results showing that higher reading scores for those offered vouchers were concentrated in the first (2004) cohort, which did not include elementary-grade students and had more private school places available to it. The effect was also focused on students scoring higher on the baseline test, on those who had not attended the most troubled D.C. public schools, and on female students. Further, some of the most interesting results of the study were related to student choice behavior rather than increases in test scores. For example, one-fourth of voucher recipients never used their vouchers, many used them only part of the time, and almost all switched schools at least once. The report could have done far better in analyzing the results of the experiment by presenting them in a more nuanced fashion that focused on the possibility of varied effects with different populations and in different contexts and discussed the limitations of the results for more generalized large-scale applications.
I. INTRODUCTION

Beginning with Milton Friedman’s 1955 essay,1 many political conservatives have promoted school vouchers as a primary solution to America’s educational problems. Friedman claimed that government funding (through vouchers) for families to send their children to private school would generate two big payoffs: lower-priced education through competition and greater parent satisfaction through choice. However, the underlying message went beyond this: private schooling was simply more efficient than public—it could produce the same or more learning among students at lower cost than publicly-run education.2

No U.S. state has yet voted for universal vouchers—that is, for programs in which every family could get public funding to send their children to private school.3 But targeted voucher programs, where vouchers are only available for low-income families, have been implemented in Milwaukee, Cleveland, and Washington, D.C. A major barrier to voucher plans was removed in 2002 when the U.S. Supreme Court upheld the use of vouchers in Cleveland’s religious schools.4 In all of the extant voucher plans, pupils attend private religious schools.

With the rise of charter schools, vouchers have gradually drifted into the background of the school choice movement, for three reasons: first, charters are in most states at least partially subject to public (state or school district) regulation, hence more acceptable politically to taxpayers; second, unlike vouchers, charters generally receive approximately the same funding as traditional public schools, hence attract more entrepreneurs willing to supply schooling in the alternative school market; and third, charters satisfy conservatives’ criteria that they compete with the public system and are not directly publicly managed. Nevertheless, in keeping with Friedman’s original views on private versus public, voucher proponents continue to fight for vouchers as an instrument that provides access to private, unregulated education.5

This is the political setting for the D.C. voucher plan, launched in 2004 by the Republican-dominated Congress as the District of Columbia School Choice Incentive Act of 2003.6 The Act established the first federally funded private school voucher program in the United States. Congress mandated that the plan be evaluated each year, and the current third-year report (March, 2009) follows the first evaluation report in June 2007, and the second report in June 2008.7 All three reports use the same methodology.

II. THE REPORT’S METHODS, RATIONALE, FINDINGS AND CONCLUSIONS

Since its inception, five cohorts of students have participated in the D.C. Opportunity Scholarship Program (OSP), as it is officially called. Each spring, beginning in 2004, families applied for a private school scholarship (voucher). OSP prioritized poor and near-poor families attending public schools, particularly those who attended “problem” public schools designated “in need of improvement” (SINI schools) in accordance with the No Child Left Behind law. Eligible children (poor or near-poor who attended public schools or those just entering kindergarten) were baseline tested and families filled out a baseline questionnaire.
Depending on the number of places available in one of the approximately 66 private schools agreeing (in 2004) to accept OSP students, that number of applicants was randomly selected from the eligible applicant pool to receive a scholarship (voucher) of $7,500. The rest did not receive a voucher and were designated the control group.8

This randomized assignment method generally provides a strong test of a causal relation between an intervention and differences in outcomes between treated and control groups. For all its strengths, however, the method has its problems.9 In particular, participants know which group they are in, and the results may be context-sensitive since the experiment necessarily takes place in a defined set of conditions. Further, as the experiment proceeds beyond a year or two, it is increasingly difficult to induce the members of the original sample to take annual tests for the evaluation. This D.C. study has both the strengths and weaknesses of such experimental designs.

In practice, most students in the D.C. experiment used the vouchers in a relatively small number of private schools, and most of these were in religious schools. Specifically, 82% of voucher users attended one of the faith-based private schools participating in the program (faith-based schools generally charge lower tuition so the voucher was more likely to cover the entire cost of attendance), and 59% attended one of the 22 Catholic schools. As discussed in more detail later in this review, not all students who received vouchers used them, for various reasons, including not being able to find a place in a suitable private school, and a very high percentage (90%) switched schools in the first year attending a private school. (See Figures 2-3 and 2-4 of the report for a clear picture of voucher use and switching. It should be noted that a high fraction of these switchers moved to a different private school, and an unspecified number also moved between schools as part of regular advancement from, e.g., middle school to high school).

The evaluation of the OSP consisted of comparing the achievement scores in reading and mathematics of the randomly selected voucher recipients in Cohorts 1 and 2 (the first two years of the program) with the Cohort 1 and 2 control groups one year, two years, and three years after they applied to the program. For the three evaluations, Cohort 1 was retested in spring 2005, 2006, and 2007, and Cohort 2 was retested in spring of 2006, 2007, and 2008.

As explained below, members of the two Cohorts had notable differences in their school placement and their outcomes. At a minimum, the Cohort 1 sample differs from the Cohort 2 sample in that only a subsection of the Cohort 1 students could be included in the experiment. When Cohort 1 students applied in 2004, all eligible students of K-5 age could be placed in a private school. (This was an “abnormal” situation and was not repeated in 2005.) Hence all eligible K-5 students (851 students of 1,343 eligible applicants were in these earlier grades) in Cohort 1 were offered vouchers, leaving none for the control group and meaning that the students were not included in the evaluation. The rest of the eligible students in Cohort 1, who were in grades 6 and higher (the “impact” group), were divided into voucher recipients (299) and non-recipients (193), the latter being the control group. According to Patrick Wolf, the lead author of the evaluation, the vast majority of the Cohort 1 sample consisted of middle school students, although this is not transparent in the results presented.10

The Cohort 2 (spring 2005) impact group (those offered a voucher, as well as the
control group) was larger than Cohort 1, for two reasons: first, a higher fraction of students was eligible among those who applied, creating a larger pool for the randomized impact group, and second, a smaller fraction of the eligible students could be placed in private schools, so the number who ended up in the randomized sample was much larger—1,088 who were offered vouchers and 728 who were not. Cohort 2 students were spread out over K-5, middle school, and high school. Thus, 78% of voucher recipients and 79% of the control group were in Cohort 2. For the reason set forth above, all primary school (K-5) students in the sample ultimately analyzed were also in Cohort 2. Most non-recipients in the combined Cohort 1 and 2 sample returned to their assigned public schools, but some returned to or switched to charters or managed to go to private school despite not being offered a voucher (Table 2-7).

Because students were randomly assigned to receive a voucher (the treatment group) or not receive a voucher (the control group), the treatment analyzed in the report is the offer of a voucher (intent to treat), not its use. The randomizing process could not control who actually used a voucher. Therefore, the third-year evaluation estimates the difference in achievement scores at the end of the third year between those who were offered a voucher and those who were not. The comparison in this third-year report is based on the spring tests in the third year that each of the two cohorts was in the program. This estimate is then adjusted for the percentage of voucher users on the assumption that the achievement scores of those students who did not use the voucher would not have been affected by going to private schools (although some did so on their own).11

The estimates from the overall impact sample show that in the third year after applying to the program, the treatment group—those students who were offered vouchers—scored 4.46 points higher than the control group in reading and less than a point higher in math (Table 3-2). This is the total difference on the third-year test. The reading score difference is statistically significant, but the math difference is not. When adjusted for the percentage of those remaining in the third-year sample who actually used the offered voucher to attend a private school, the reading score difference is 5.27 points (an effect size of 0.15 standard deviations). The adjusted math score difference is less than a point and is not statistically significant. In the previous two evaluations (years 1 and 2), the results showed no significant differences between treatment and control groups, either in reading or math.

In addition, the evaluation estimates differences between the voucher offer group and the control group in parent and student perceptions regarding school safety and in satisfaction with their school. Voucher parents rate their children’s schools significantly higher on safety and are significantly more satisfied with their children’s school. However, there is no significant difference between voucher and control group students’ views on either perceived safety or school satisfaction.

The evaluation also tests for differences across certain subgroups in the sample, which allows for exploration for the particular source of any benefits. For example, the effect on reading achievement score of being offered a voucher is much larger in Cohort 1 than in Cohort 2, where it is positive (about 3 points) but not significantly different from zero (Table 3-3). This is important, because, as noted above, none of the Cohort 1 students in the impact
group attended primary schools, and almost all attended middle schools. The Cohort 1 students were also able to select from a less constrained group of private schools. In the words of the report, “Cohort 1 students faced a different set of participating schools and fewer slot constraints in those schools than did cohort 2 students, conditions that could generate variance in program impacts” (p. 33, emphasis added). This suggests that much of the reading achievement benefit reported in the third-year evaluation was for a treatment of middle school students who were able to select from among a relatively larger group of (religious) private schools.

The effect is also much larger for students who never attended a SINI school (a D.C. school that has been designated as “needing improvement”). Congress designated that students from SINI schools should be most targeted by the OSP. Yet, there was no significant effect on their scores from being offered a voucher. The authors imply that not too much should be made of the lack of a voucher effect on those from SINI schools because all students eligible for the vouchers were from families within 185% of the poverty line. Yet, in terms of the Congressional mandate, vouchers did not provide an academic benefit to students from the schools most needing improvement.

Other subgroups showed significant differences between those offered vouchers and those not. Females who received voucher offers had a statistically significant larger third-year score than females in the control group, but males showed no difference. Students in the treatment group who entered the Program in the higher two-thirds of the applicant test-score performance distribution—averaging a 43 National Percentile Rank in reading at baseline—scored significantly higher than the higher initial scoring control group, but that was not true for the treatment group that had lower initial scores. As well, students in the treatment group who attended school in K-8 did significantly better than the control group, but that was not true for treatment students who attended high school.

Since the treated students in Cohort 1 were concentrated in middle schools and the effect on their reading score was significantly higher than for treated students in Cohort 2, it is highly likely that the significant effect in K-8 is really only a 6-8 effect, and that the difference between the treatment and control groups in K-5 (entirely Cohort 2) was not statistically significant. Thus, the subgroup analyses suggest that the reading benefits after three years were larger for females, students with relatively higher initial reading achievement scores, and middle school students.

III. REVIEW OF THE VALIDITY OF THE REPORT’S FINDINGS AND CONCLUSIONS

What does this evaluation tell us about the potential effect of offering vouchers in D.C., or, for that matter, in any other low-income urban school district?

At one level, and the one the reports’ authors emphasize, sending low-income students to existing, predominantly religious (and even predominantly Catholic), small (average size, 265 students) private schools with small class sizes (average student-teacher ratio, 10.3 students) can modestly increase these students’ achievement (in reading but not mathematics) and result in greater parent satisfaction with their children’s school.

At another level, the study implicitly reveals a lot about such programs that the authors
faithfully report but do not emphasize. First, the OSP is necessarily small, not because of lack of scholarship funding, but because of the number of places made available by participating schools. It is not possible to tell exactly how many places the 102 private schools in the District of Columbia made available to the OSP, but we do know that initially 66 schools participated and that the number in 2007-2008 had declined to 60. We also know that the number of scholarship awardees declined from 1,366 in 2004 to an average of 350 awardees in 2006-2007 (Table 1). One of the main reasons for the decline is that once the first two cohorts had been placed, it became progressively more difficult to find places for new awardees. There were too few private school slots. Getting new scholarship recipients into participating schools depended increasingly on OSP students dropping out of the program or leaving a private primary school for middle school or a middle school for high school.

This means that the total of 1,700 voucher users in fall of 2008 represents the approximate capacity of the program. Total enrollment in D.C. public schools was about 49,000 students in 2007-2008, and in D.C. charter schools, about 22,000, for a total of about 70,000. So the 1,700 places represent 2.4% of D.C. enrollment and could not be expected to increase significantly even if Congressional funding were maintained.

Second, the OSP results suggest great variation in voucher usage among recipients (Figure 2-3) and a high degree of school switching (this includes “natural” switching from primary to middle to high school) among all students in the sample, and even greater among voucher recipients (Figure 2-4). Of the 1,387 voucher recipients in the two cohorts, 1,041 used the voucher for at least a short time, 750 used it for the first two years, and only 568 used it all three years. Some of the drop in use from the second to the third year could have occurred because a school level was completed and an appropriate school at the next level was not available, but these still represent low utilization rates for an attractive opportunity to leave conventional public schools. Further, only 3% of the treatment group and 15% of the control group never switched schools over the three-year period. By the end of the first year, 90% of the treatment and 58% of the control group switched schools. Even assuming that some of those were switching because they moved from one level to the next, these rates are still very high. Clearly, getting a voucher offer or even using the voucher that first year did not mean that the user stayed in his or her initial school.

Third, the positive effect of attending private school seems to have been concentrated in the subgroup of students who were more academically adept before they were offered the voucher. True, all of these “higher-scoring” students were relatively low-income and averaged considerably below the 50th national percentile in reading, but attending private school does appear to have worked best for the abler students, to the extent that it raised reading scores but not math scores. This is not a surprising result, in the sense that given the degree of switching schools, less-able students attending private schools may have switched more often than more able students and more often than the less able students in the control group. Less-able students in the treatment group may also have used the voucher more sporadically. We have no way of knowing about such differences from the way the data are presented.

Fourth, the considerably larger treatment effects on reading (and math) achievement
in Cohort 1 suggest that the overall positive results for the treatment group in Cohorts 1 and 2 combined are being driven largely by the difference between the treatment and control group in Cohort 1. Since randomized assignment took place within each of these two cohorts, it should be relatively easy to compare how the two cohorts differ. (Such an analysis is not presented in the evaluation.) For all we know, the Cohort 1 students also may have averaged higher scores on the baseline test than those in Cohort 2.

There is also a reasonable possibility that the first cohort was less typically “treated” than Cohort 2. We know that Cohort 1 students were concentrated in middle schools and the treatment group had a much better choice of places in private schools. But this situation was described as “abnormal” and therefore provides weak support for generalizations to normal situations. Given the apparently large difference in treatment effect between the two cohorts, the report should have included, in Chapter 3 (“Impacts”), a discussion of possible differences between them. Perhaps it makes sense that middle school is the level where attending a private school would have the largest effect on reading scores. It would have been useful to test whether Cohort 2 middle school voucher recipients also had a statistically significant higher reading score.

Finally, the study highlights some of the major problems of randomized trials. The first is context, which is brought out in a small way by the different results for Cohorts 1 and 2. If the particular treatment conditions for Cohort 1 students are what carried the day, vouchers per se may be less the relevant “treatment” than the offer of a voucher combined with particular recipients and particular use. The authors might have emphasized that vouchers for certain groups in certain levels of school (the abler among low-income students attending middle schools, for example) result in a modest treatment effect.

The second problem for randomized trials is loss from the original sample. The loss is very large in both the treatment and control groups. The authors of the report are well aware of the problem and try to deal with it, but losses of this magnitude (32%) always cast doubt on whether the patterns of loss in treatment and control groups do not differ in some way that might bias the estimated effect of the treatment.

IV. USEFULNESS OF THE REPORT FOR GUIDANCE OF POLICY AND PRACTICE

Should this study change the way we view vouchers? For voucher proponents, a randomized field trial result that shows a third-year effect size of 0.15 standard deviations (even if in only one of two subjects) is obviously meaningful, as suggested by the tone of the report and editorials in the Wall Street Journal and Washington Post. The argument in the Post was that since the students receiving vouchers were academically a bit better (or at least no worse) off than those who did not and that parents were more satisfied, the program should continue. Setting aside for a moment the methodological and other concerns raised above, that is a fair argument in terms of saving a particular program that is no more costly per student than is spent in D.C. public schools.

However, it is not a particularly good argument for those seeking to make large improvements in academic achievement for low-achieving, low-income students. After many years and much voucher research—some randomized trials and others
comparing voucher students with similar students in public schools—the results fail to make the case. They strongly suggest that voucher students have either not done significantly better, or that the differences have been relatively small and inconsistent across academic subjects and groups of students, as in the D.C. Opportunity Scholarship Program.\textsuperscript{17} After all, if the most we can hope for from an intervention is that students attending middle school increase their reading score one-sixth of a standard deviation in three years, but their reading does not continue to get better in high school (and their math scores don't improve at all), that is not going to help very much.

The modest result in D.C. does not appear to be enough to convince voucher opponents or even neutral policymakers that vouchers should be an important component of a broader attack on closing the achievement gap between low-income black students and middle class whites and Asian Americans. Given the political cost to the Obama administration of abandoning its campaign position on vouchers, the difference in achievement estimated in this report was simply not large enough to support extending the program beyond those students currently in it.\textsuperscript{18} This is especially true in D.C., with its already broad and increasing availability of charter schools.\textsuperscript{19}

Further, since in D.C. only a limited number of voucher students were placed, and those placements were mostly in existing Catholic schools of proven quality, and since the control group students were presumptively largely enrolled in D.C. public schools, which do not have a strong reputation for quality, the take away of all three years of evaluations could easily be that the results were disappointing. If students in a boutique voucher program cannot show big gains, we would not expect much from bigger programs such as Milwaukee’s—a more typical example of what a large-scale voucher program would look like—where most of the 18,000 voucher students attend private schools started up to take advantage of vouchers, and where a much smaller percentage attends the older, established Catholic schools. A recent second-year (of five years) study (not a randomized trial) of relative gains among similar social class/race students showed no significant difference in one-year gains (2007-08 versus 2006-07) between Milwaukee voucher students attending private schools and students attending public schools.\textsuperscript{20} The research is ongoing, so we shall see what the future brings. But the D.C. results should not generate great expectations.
Notes and References


3 However, tax credit voucher policies are on the increase. These policies provide state reimbursement, through tax credits, when a taxpayer donates to a non-profit organization which bundles the donations and issues vouchers to parents to pay for private schooling. Tax credit vouchers have been adopted in six states, including Arizona and Georgia. See:


5 Most states have some regulations for private schools. As the Supreme Court acknowledged as far back as 1925, in Pierce v. Society of Sisters, 268 U.S. 510, 534, states have the right to regulate such schools including personnel qualifications, textbooks, and other details. States do not use much of their authority because of political opposition by a small and well-organized and vocal constituency of private school advocates.


8 The amount of the voucher was generous by voucher standards, even though considerably less than spending per student in D.C. public schools. Since the private schools involved provided a lower level of services such as special education for more severe disability cases and provision for English language learners (see Table 4-1), their costs per student also tended to be lower than the average D.C. public schools.


10 E-mail correspondence with P. Wolf, May 5, 2009

11 The Intent to Treat effect is divided by the percentage of students in the treatment group using a voucher in the third year. Appendix 5 (p. A-30), reports that “of the scholarship awardees in the impact sample that provided year 3 outcome test scores, 86% had used a scholarship for all or part of the 3 years after the random assignment.” In the total sample (includes those that provided test scores or not in year 3) the
percentage of those offered a voucher and using it was 75%. It is difficult to know whether the higher percentage of voucher users in the treatment group test takers in year 3 is correlated with some unobservables that would make them do better on the reading test than those who did not take the test.

Nevertheless, in one of the odder reported results of the study, the mean student-teacher ratio for the treatment and control groups was essentially the same (13 students), and only the school size differed substantially (Table 4-1).

One explanation may be that Washington is a city with many charter schools providing an alternative to both public and private schools.

Note that even the abler students had rather low (43rd national percentile) baseline scores.

The claim that the large gain in Cohort 1 is a “false discovery” (p. 41 and Appendix Table B-1) does not change the likelihood that without Cohort 1, there would have been no positive effect on reading scores in the overall sample.


Yet, there is no evidence that students in D.C. charter schools have greater gains in test scores than similar students in D.C. public schools (see Schneider, M. & Buckley, J. (2007). *Charter Schools: Hope or Hype?* Princeton, NJ: Princeton University Press).