September 17, 2000
Letter to the Editor Department
New York Times

Dear Editors:

Your September 15 story, "New Doubt is Cast on Study that Backs Voucher Effects," suggests a level of disagreement among researchers that simply does not exist. All the researchers involved in evaluations of pilot programs in New York City, Dayton, and Washington, D. C., including Mathematica Policy Research (MPR), agree on the following: 1) Randomized field trials, the technique used to evaluate these three programs, are the best way to assess the effects of social interventions. 2) No positive effects of vouchers on the test scores of non-African American students were detected. 3) Test scores of African Americans switching from public to private schools for two years were higher than those who did not; these results were statistically significant. 4) The estimated effects in all three cities are not contaminated by the fact that only some of those offered the vouchers made use of them. 5) The estimated effects were unlikely to have been materially affected by the fact that not all students were tested two years later. 6) Caution is needed when drawing policy implications from a two-year evaluation. Still, issues have been raised that deserve a response.

Consistency of Results

Some have questioned the consistency of our findings. But the results from the three voucher evaluations are in fact quite similar across all three cities. In no city did we find any positive effects of vouchers on the test scores of students other than African Americans. But African Americans in New York who switched from public to private schools for two years scored 4 percentile points higher than the control group. In Dayton and D. C., they scored 6 and 9 percentile points higher respectively. The results in all three cities are statistically significant. On average, these impacts are about 6 percentile points, a moderately large effect.

Some of the researchers involved in the New York evaluation feel that the results there are not sufficiently consistent across grade levels to warrant the conclusion that voucher impacts had been detected. Although David Myers, senior fellow at MPR, agrees that statistically significant, positive effects of 4 percentile points on the tests scores of all African Americans were observed in the New York evaluation, he points out that, when examined by grade level, statistically significant effects in New York are limited to sixth graders and, as a result, he concluded, according to your reporter, that there was "no impact" in New York City. (Myers was not involved in the D. C. and Dayton evaluations.)
Fluctuations in findings often occur when one examines grade specific information. For this reason, noted education statistician, Anthony Bryk, together with his colleagues, have recommended that conclusions about school impacts not be drawn from "only single grade information...Judging a school by looking at only selected grades can be misleading. We would be better off, from a statistical perspective, to average across adjacent grades to develop a more stable estimate of school productivity."  

Bryk et al.’s admonition is particularly compelling when, as is the case in New York, only 50 to 75 African American students are observed in the treatment and control groups at each grade level. Under these circumstances, the fluctuations from one grade to the next are unlikely to generate what Bryk et al. call "stable estimates of school productivity." For this reason, it is premature to conclude from grade-specific information that vouchers had no impact on African Americans, especially when significant, positive effects are observed in Dayton and Washington—and in these cities positive impacts are not concentrated at any particular grade level.

**Take-Up Rate (Decliner Rate)**

Some have objected to our results on the grounds that about half the students took the voucher that was offered to them (the takers) and about half did not (the decliners). The percentage of students who took the voucher in each city is clearly stated in our reports. Unfortunately, your reporter overlooked these findings and instead accused us of suppressing this information.

As we have discussed in our reports, takers and decliners differed in a number of respects. Most notably, takers had higher family incomes in New York and D. C. but lower incomes in Dayton. The New York and D. C. findings are not surprising, given that the voucher awards did not cover all the costs of a private education. These additional costs were the reason most frequently given by families for not using the voucher. Presumably, take-up rates would rise if the monetary value of vouchers were increased.

Nevertheless, the fact that takers and decliners differ in income and other respects does not bias our estimates of the impacts of vouchers on test scores. Both MPR staff and we agree that the sophisticated but widely used instrumental-variable technique that we employ effectively adjusts for these differences. This analytical technique takes advantage of the fact that vouchers were offered at random. It was first used in medical research, is now commonplace in econometric studies, and was employed by Alan Krueger in his study of the effects of class size on student performance in Tennessee, a study praised by many of the same people who have criticized our report.

---


Testing Participation Rates

Others have pointed out that not everyone in the test and control groups continued to participate in the evaluation two years later. This problem, which is encountered by virtually all evaluations of social interventions, is a valid concern. We did our best to locate and persuade as many families as possible to continue to participate in the evaluation, whether or not they had received a voucher. Still, for a variety of reasons, substantial numbers of students were not tested at the end of the second year.

Both MPR and we are reasonably confident that this problem does not undermine the integrity of our findings. We obtained the test scores and background characteristics of virtually every student involved in the study at baseline, before they were randomly assigned to treatment or control groups. These data reveal only minor differences between the second-year participants and non-participants in all three cities (as detailed in our reports). To account for the modest differences we did observe, we weighted the data, a technique that MPR endorses.

In a randomized field trial, it is desirable to have similar response rates for test and control groups. If response rates differ noticeably, it is possible that the two groups participating in the study will no longer be comparable. Fortunately, in Dayton and D. C., the response rates for the test and control groups were essentially the same. Only in New York City did the students in the control group participate at a lower rate than the students offered a voucher—here the difference was 7 percentage points.

Need for Caution

As we emphasized in our report, which was presented and critiqued at the recent meeting of the American Political Science Association, one needs to exercise caution when drawing policy conclusions from our findings. These are only two year results from fairly small pilot programs. Over the long run, results may become positive for all ethnic groups, or the effects of the program on African American students may dissipate altogether. And larger voucher programs may have quite different effects. Still, the weight of our evidence challenges both voucher advocates and their critics: Positive impacts were consistently observed for African Americans but not for any one else.

William G. Howell, University of Wisconsin
Patrick J. Wolf, Georgetown University
Paul E. Peterson, Harvard University
David E. Campbell, Harvard University