Voucher and Class-Size Research

By Alex Molnar and Charles Achilles
Education Week, Oct. 25, 2000

The academic benefits of educational vouchers were blockbuster back-to-school news this autumn. The report that generated all the media attention, "Test-Score Effects of School Vouchers in Dayton, Ohio; New York City; and Washington, D.C.: Evidence From Randomized Field Trials," was authored by a team of researchers, including Paul E. Peterson of Harvard University. Shortly after the voucher report was released, former U.S. Secretary of Labor Robert Reich wrote an opinion piece in *The Wall Street Journal* making a case for what he called "progressive vouchers." Referring to the latest voucher study, Mr. Reich asserted: "Evidence mounts that vouchers do work for kids who use them."

If by "work" Mr. Reich means that voucher programs increase academic achievement, perhaps he should review the evidence with more care. Virtually all of the data on the achievement impact of publicly funded voucher programs come from the Milwaukee Parental Choice Program and the Cleveland Scholarship and Tutoring Program. Although the Cleveland evaluation is ongoing, the Milwaukee data are now 5 years old. Moreover, the 1995 legislation that expanded the Milwaukee program to include religious schools also removed its evaluation component, so no new data will be available anytime soon. Therefore, most of the evidence about the impact of voucher programs will, for the foreseeable future, come from the privately funded voucher programs, such as those in Dayton, New York City, and Washington.

In the August 2000 report referred to by Mr. Reich, the Peterson team presented the results from Dayton, New York, and Washington averaged across grade levels. The combined averaged results from all three cities are also presented. Since averaged results tend to conceal inconsistent findings, they may make the achievement impact reported appear more generalized than it is.

In this case, even the averaged results are inconsistent. Considering the results that are significant at the conventional .05 level yields the following information: According to the Peterson team, simply being offered a voucher makes a significant difference in reading but not math achievement for African-Americans in New York City in year two. In Dayton, the offer produces no significant result in either year. In Washington, the offer makes a difference in African-American test scores in reading and math in both years. Actually *switching* to a voucher school in New York produces a significant advantage in reading for African-American students in both years, and in math in year one. In Washington, there is an advantage in math in both years, and in reading in year two only. In Dayton, switching to a voucher school produced no achievement advantage in either year.

Except in Washington, for which two grade-level groups (grades 2-5 and grades 6-8) for African-American students are provided, it is not possible to tell how students in different

grades performed. The Washington results, however, appear to be driving the averaged findings for the three cities. More specifically, it seems that the averaged results are heavily influenced by African-American students in grades 2-5 in Washington.

The impression created by the averaged data was apparently troubling enough for Mr. Peterson's New York research partners at Mathematica, a highly respected private research firm, to cause them to issue a separate statement, entitled "Voucher Claims of Success Are Premature in New York City." In their statement, the Mathematica researchers reported that students offered a scholarship performed at about the same level as students in the control group. They further noted that the gains reported for African-American students in New York were largely the result of the scores of one group of students (who were in the 5th grade in 1998-99 and 6th grade in 1999-2000) and, therefore, must be interpreted with caution. The Mathematica statement pointed out that there was no achievement advantage for Hispanic students and advised that, since the gains reported were so concentrated in a single group, "one needs to be very cautious in setting policy based on the overall modest impacts on test scores."

After reviewing the major research reports associated with the academic impact of vouchers, setting aside methodological qualms, and accepting at face value all findings that achieved the conventional .05 level of significance, we offer the following advice to parents who are concerned about the academic performance of their children and who are interested in vouchers:

• In Milwaukee, enroll your child in a voucher school and wait three or four years. Your child may at that point begin to do better in math and reading (research by the Peterson group), or just in math (research by Princeton University's Cynthia Rouse), or in neither subject (research by the University of Wisconsin-Milwaukee's John Witte). A better choice would be to send your child to a Milwaukee public school with reduced class size. They outperform voucher schools (Rouse).

• In Cleveland, enroll your child in a voucher school for 4th grade language arts and maybe science, depending on the classroom to which your child is assigned. Be careful, however, to avoid newly created private schools. Students in those schools perform significantly less well than students in the Cleveland public schools in all subjects tested (research by Indiana University's Kim K. Metcalf).

• In Washington, if you are an African-American parent, enroll your child in a voucher school in grades 2-5 (Peterson).

• In New York City, if you are an African-American, enroll your child in a voucher school in 7th grade for the 2000-01 school year (Mathematica).

Even without considering the manifold methodological questions that threaten the validity of many voucher findings, the literature on vouchers and academic achievement adds up to an inconsistent gaggle of results that do not offer intelligible guidance to either parents or policymakers. Unfortunately, as Jay P. Greene previously has noted in

Education Week (“Rescuing Education Research," Commentary, April 29, 1998): "Interest groups have learned that they can successfully check research contrary to their goals by producing their own studies, no matter how lousy, to sow confusion among policymakers, journalists, and the attentive public about what to believe."

Thus, despite the preliminary character and inconsistent outcomes of voucher research, well-financed voucher advocates are now sowing confusion by claiming that the achievement evidence is clear, systematic, and compelling. Some compare the power of the research and the magnitude of the voucher findings to the results of class-size research, most often the Tennessee Student-Teacher Achievement Ratio, or STAR, experiment. Since education policy is a matter of making choices among different policy options, that contention is worth examining.

In general, we can have the most confidence in a body of research that includes studies that are of sufficient size and scope to provide educationally important and reliable results. Moreover, current work should support the findings of earlier studies, and the findings should be the subject of rigorous peer review. Using these criteria, there is no comparison between class-size and voucher research.

The STAR project, conducted in Tennessee from 1985 to 1989, was a longitudinal, randomized experiment involving 11,600 students, grades K-3. STAR was designed to determine the effect of reducing class size in grades K-3 on student achievement and social development. At the end of the experiment, STAR students were followed throughout their K-12 education careers. They continue to be monitored in their postsecondary pursuits. Thus, STAR provides a large-scale, longitudinal assessment of the impact of reducing class size in the primary grades. Initial STAR results, showing a significant achievement advantage for students in small classes, and numerous follow-up and subsidiary studies, have been published in refereed journals.

The initial STAR data have been reanalyzed and the results confirmed by the Princeton University economist Alan Krueger. In a subsequent study, he also found a significant relationship between participation in a STAR small class and the rates at which students took college-entrance exams.

Follow-up studies of STAR students in grades 4-12 have found that they outperformed students who had been in regular-size classes on tests in all subject areas tested. In fact, students who had small classes from kindergarten to grade 3 outperformed regular-class students on all tests, every year, and the gap between small and regular-class test results *increased* each year. For example, an average STAR student in a small class for grades K-3 would outperform a regular-class student at grade 4 by 6.6 months and by 8.7 months at grade 8, fully five years after leaving the small-class environment.

The magnitude of positive voucher academic effects is sometimes described as comparable to the total STAR effects, rather than the STAR effects for minority students. For voucher studies, in which the overwhelming majority of students are African-American, the relevant comparison is to the performance of STAR minority students. For

those students, the STAR effects were approximately double the total effects. Interestingly, it appears that the voucher students who perform at a higher level than students in public schools tend to attend smaller classes in smaller schools.

In addition to STAR, the positive results of class-size studies in, for example, Wisconsin; California; Fairfax County, Va.; and Burke County, N.C., complement the findings of powerful early-intervention studies. The Perry Preschool study followed 123 youngsters, and the Abecedarian study followed 109, from preschool into adulthood. The treatment-group students in these studies have consistently demonstrated better educational and social outcomes, such as employment, than their control-group peers. When the results of these early-intervention projects are considered in light of the STAR data on the enduring academic impact of small classes in grades K-3, all of the findings are strengthened, and policy implications can be drawn with considerable confidence.

It is now possible to say with a high degree of assurance that attendance in a small class is not a one-time inoculation that assures academic success. For long-term, continuing benefits, children should participate in small classes in the primary grades or preschool. The small-class experience must be intense-all day, every day, for a sufficient duration (that appears to be at least four years). When this combination of early intervention, intense treatment, and sufficient duration occurs, academic and social benefits continue to increase throughout schooling and extend into young adulthood.

It is no surprise that the recent RAND Corp. study of American education, "Improving Student Achievement," found strong associations between student achievement and both higher public prekindergarten participation and reduced pupil-teacher ratios in the lower grades.

Some voucher advocates have taken to supporting their position by asserting that no one has demonstrated that vouchers harm public schools. In our view, this is not good enough. Good public policy should be constructed using the most powerful data available to actually help the public schools provide a high-quality education to the students they serve. When it comes to the evidence on vouchers and the evidence on class size, the policy choice is clear.

Alex Molnar is a professor of education and the director of the Center for Education, Research, Analysis, and Innovation at the University of Wisconsin-Milwaukee. He is a principal investigator of Wisconsin's Student Achievement Guarantee in Education (SAGE) class-size-reduction program. Charles Achilles is a professor of education at Eastern Michigan University and was a principal investigator of the Tennessee Student Teacher Achievement Ratio (STAR) class-size experiment and numerous other class-size studies.