VOUCHERS AND SCHOOL CHOICE: THE EVIDENCE

Greg Forster

In the USA, states are increasingly adopting school choice policies, and a large body of empirical research using especially high-quality methods has studied their effects. It consistently finds that school choice improves academic outcomes for those who use it, improves outcomes at nearby government schools, and has other positive effects.

Introduction

One of the perennial problems of politics is the difficulty of getting people to base their opinions on a systematic analysis of real-world evidence rather than myths and anecdotes. Nowhere is this problem more severe than in the field of education policy, where trade unions protect their gravy train by inculcating myths about the virtues of a government monopoly over education and the evils that market forces would unleash if they were applied to schools.

The good news is that a large and growing body of empirical evidence is available to counteract the prevalence of education myths. And the largest and most scientifically rigorous body of evidence concerns school choice. The research consistently finds that school choice policies have positive effects across a wide array of outcomes.

School choice is controversial because people have different expectations about how market forces will impact on education. Opponents claim that if we move from the current education monopoly to an education market, the wealthy may get better services, but the poor would get worse services. Advocates respond that markets would provide all students with better services than the current monopoly system gives them.

There is a large body of empirical evidence that ought to inform this debate. In the USA, more and more states have adopted school choice policies; there are now 21 school choice programmes in 13 states plus the District of Columbia. Almost 190,000 US students currently attend privately owned schools (including both religious and non-religious schools) using government funds, through either voucher programmes or 'tax-credit scholarship' programmes that accomplish the same purpose but are funded through the tax code.

Scientific studies have examined how these programmes have an impact on every aspect of education from academic results to segregation and civic values. Unfortunately, the debate over school choice too often is dominated by myths and anecdotes rather than the evidence this empirical research provides.

When reviewing this research, it is important to bear in mind that privately owned schools in the USA enjoy much greater autonomy than is the case in most European nations. State laws regulate privately owned schools for health and safety, and require them to teach a standard set of academic subjects. However, the schools are mostly free to set their own curricula and follow their own educational philosophies and methods. And, contrary to what some have feared would happen, school choice programmes have not resulted in significantly increased regulation of privately owned schools.

The autonomy of privately owned schools is a crucial element in school choice, because a choice among options that have all been standardised by heavy-handed government regulation would not be a real choice at all. Advocates of education reform in other countries should bear this in mind when considering what reforms are most needed in their own systems.

Why methods matter

It is especially important to note the high scientific quality of the research on school choice. Usually, it is very difficult to study the effects of education policy properly, because student outcomes are affected by so many different influences – including demographic factors (income, race, family structure, etc.),

school type (government or privately owned) and intangibles such as the level of enthusiasm parents and teachers invest in a child's education. The job of empirical science is to disentangle the influence exercised by each of these factors as well as can be done with the available evidence.

When it comes to comparing government and privately owned schools, the problem is compounded by selectivity. In the USA, students attending privately owned schools are there because their families chose to make a financial sacrifice to put them there. In addition, some privately owned schools are selective to some degree in admitting students. Thus, any observable differences between students in government and privately owned schools may be due either to differences in the schools or to parental and school selectivity.

Making things worse, many of the factors that affect student outcomes are not measurable. The enthusiasm that parents and teachers invest in a student's education has an important impact on the student's outcomes, but we have no way to measure it. Thus, such factors cannot be controlled for statistically.

This is where the scientific quality of the evidence comes in. A study that uses good methods can overcome these problems and provide reliable information about what is influencing student outcomes. But if scientific procedures are not rigorously followed, we can come to the wrong conclusions about what factors cause what outcomes. A poor- or even mediocre-quality study is more likely to: falsely attribute causal power to a factor that doesn't really matter; falsely attribute no causal power to a factor that does matter; or falsely attribute one type of influence to a factor that actually exerts a different type of influence.

The gold standard for empirical science is the method known as 'random assignment'. In this method, subjects are randomly divided into two groups: a group that will receive the treatment being studied (such as school choice) and a control group that will not receive it. Because the two groups are separated only by random assignment, they are likely to be very similar in every respect other than the treatment being studied. Thus, if the two groups have different outcomes, researchers can attribute that difference, with a high degree of certainty, to the treatment.

Random assignment studies are very rare in social policy. We do not usually have the opportunity to divide populations by random lottery and apply different policies to them. However, school choice has provided researchers with valuable opportunities to conduct random assignment research because school choice programmes often are oversubscribed. When too many people apply to participate in a school choice programme, a random lottery often is used to determine which students will be invited to participate. This creates a naturally occurring random assignment situation – applicants who are invited to participate as a result of the lottery are the treatment group and applicants who are not invited are the control group. Both groups are made up of students whose parents applied to participate in the programme; they are separated only by whether their applications were accepted as the result of a random lottery.

When a significant body of random-assignment research exists, its findings should take precedence over the findings of other types of studies. No method is as good as random assignment at disentangling the influence of a treatment from the influence of other factors.

However, this is not to say that random-assignment research is the only kind worth considering. It may be the best kind of research, but where random-assignment research cannot be conducted, other kinds of research are well worth conducting.

The next best research method is to track year-to-year changes in outcomes for individual students. Although it is not as good as random assignment, this is still a very good method, and its results are widely regarded as being high in scientific quality. Tracking individual students over time removes from the analysis most, though not all, of the influence of unmeasured factors. If a student is advantaged in a way that is not measurable, that advantage will typically be present in the student's outcomes for both year one and year two of the study; thus the change in outcomes between year one and year two will mostly be due to other factors – though unmeasured factors will still exert some influence on the level of year-to-year change.

If it is not possible to track individual students, good research still can be done by tracking year-to-year changes in individual schools. The unmeasured advantages of the students in a given school can reasonably be expected to be similar from year to year. If a school has highly advantaged students in 2006, it probably will still have highly advantaged students in 2007. Mobility among the student population will create some change in student characteristics from year to year, but not so much that we cannot learn from school-level studies.

When individual schools cannot be tracked, some other methods are scientifically acceptable, but should be accepted with a lower level of confidence. The significance we attribute to the results of a given study should depend on the method it uses and the nature of the question being addressed.

Effect of vouchers on academic achievement

Ten analyses of school voucher programmes have used random-assignment methods. Since random assignment provides top-quality evidence that removes the effects of selection from study results, this evidence ought to take precedence over studies using other methods.

Of the ten studies available, eight find that students using school vouchers had higher levels of academic achievement than students who applied for vouchers but lost a random lottery and did not receive them. The other two studies also found positive results for vouchers, but in these two studies the results failed to achieve statistical significance, meaning that we cannot be at least 95% certain that the positive results are real and not the result of a fluke. In both of these studies, as we will see below, there is strong evidence that the failure to achieve significance can be explained by circumstantial factors. Overall, this constitutes an extremely strong body of evidence in favour of school vouchers (see Forster, 2007).

Like all studies, these random-assignment studies are limited. They do not tell us everything. For example, because of the high level of mobility that prevails among the disadvantaged populations these programmes were serving, the studies are not able to track students over very long periods of time; the longest period of analysis is four years. Moreover,

two of the studies examine a programme in Charlotte, North Carolina, for which baseline achievement data are not available. These baseline data would tell us how the students in the treatment and control groups were performing before they entered the programme, which would provide some additional statistical certainty about the results, and also allow us to confirm empirically that the treatment and control groups started out similar in their characteristics (as they ought to be if the random lottery were properly carried out).

In a set of three studies performed by William Howell and Paul Peterson (2002) of Harvard University, the positive results for privately funded voucher programmes in three cities were statistically significant for black students but not for other student groups or for the whole student population. As the authors point out, since black students are the most consistently underserved by government schools, they stand to gain the most from being offered a choice – and thus their improvements are easier to discern statistically.

Interestingly, this limitation on the positive results for vouchers did not recur in a later re-analysis of the results of Howell and Peterson's study in New York, conducted by a team led by John Barnard (2003) and including researchers from Harvard, Columbia and Johns Hopkins Universities. The re-analysis found statistically significant positive results for vouchers among all students, not just black students.

These limitations in the random-assignment research on vouchers would be more serious if we did not have the same positive finding repeated over so many studies. If the two Charlotte studies were the only two studies available, we might have lingering doubts about whether there had been some problem in the random sampling. If the Howell and Peterson studies were all we had, we might wonder whether vouchers helped all students or only the most disadvantaged ones. But the six studies that do have baseline data should allay our concerns about the two that do not, and the five studies that do find positive results for all students should allay our concerns about the three that do not. To ignore the results of the majority of studies on grounds that the remaining minority of studies suffer from limitations is not a rational approach to the evidence.

Of the two studies that did not find significant results, one – a study of the new voucher programme in Washington DC by a team of researchers led by Patrick Wolf of the University of Arkansas (Wolf *et al.*, 2007) – is ongoing. In this study, voucher students had higher test scores, but the results did not achieve 95% certainty, the conventional cut off for considering results 'significant'. In mathematics, the results achieved 93% certainty – just barely missing the cut off. The reading results are much less certain than the maths results, as is often the case in education studies.

The lack of statistical certainty may be due to the study having only a year's worth of data so far. In the eight studies discussed above, five found statistically significant results in the first year, but three did not; they required more years of data to achieve significance. Given that the maths scores came so close to achieving significance in the first year, we ought to wait for future years of data before pronouncing a verdict on the effectiveness of the DC voucher programme.

Even the result that came in at 93% certainty should not be dismissed. Placing the cut off for statistical certainty at 95% is a longstanding conventional practice, just like placing the cut off for drivers' licences at age 16 in the USA. But it is essentially arbitrary. There is nothing magical about the difference between 94.9% certainty and 95.1% certainty, just as there is no particular reason to think that teenagers miraculously become responsible enough to drive at midnight on their 16th birthdays. Scientists generally recognise this fact; many of them report results as 'moderately significant' if they're at least 90% certain – which the DC results are. Obviously we should respect the fact that 93% is not the same thing as 95%. But Moses did not come down from Mount Sinai with stone tablets saying, 'Thou shalt not consider results significant unless they are 95% certain'. It would be wrong to dismiss this moderately certain positive finding because of an arbitrary cut off point.

The remaining random assignment study, conducted by Alan Krueger and Pei Zhu of Princeton University (2004), deserves separate discussion. The study is one of two re-analyses of the data from Howell and Peterson's previous random assignment study of a voucher programme in New York. Krueger and Zhu's re-analysis found that voucher students had higher achievement levels than the control group, just as in the original analysis; however, in their re-analysis the results failed to achieve statistical significance. (The other re-analysis was the one discussed above, which found statistically significant positive results for all voucher students.)

Other researchers have identified serious violations of sound scientific procedure in Krueger and Zhu's study. The original analysis used the race of each student's mother to classify students by race, which is the method used by the US Census and by most scientific research. Krueger and Zhu used racial identification from both mothers *and* fathers, a method that does not reflect the way most students really identify themselves by race and that is not generally used. Responsible scientists try to avoid making up their own new definitions of variables whenever they can because the opportunity to bias one's results by changing the definition of the variables is too great.

Worse, Krueger and Zhu applied their new definition of race to black students differently from how they applied it to other students. They classified multi-racial students with a black father as black, but classified multi-racial students with fathers of other races according to the race of the custodial parent. This selective application of the new definition of race calls into question the validity of its use. Krueger and Zhu also added to the dataset new students for whom information was missing, reducing the quality of the study's data.

Most importantly, Howell and Peterson (2004) have shown that Krueger and Zhu were highly selective in their choice of statistical models. Howell and Peterson analysed the data using 120 different statistical models and reported that all 120 find positive voucher effects, 108 of them finding statistically significant positive effects. In other words, it wasn't enough for Krueger and Zhu to use the wrong model – they had to use just the 'right' wrong model to prevent the positive results for vouchers from being statistically significant.

Unfortunately, deviation from legitimate scientific methods is not uncommon in research on education. Given the persistent problem of bad research on education policy, the existence of a large body of top-quality random assignment

studies is a great blessing. These studies provide a scientifically solid standard for evaluating school choice policies.

Effect on government schools

Perhaps the most important concern about school choice is the effect it has on government schools. Many people acknowledge that school choice helps the students who use it, but are worried that it will make government schools worse by draining money or by 'creaming' the best students.

However, the evidence on the real-world effect of existing school choice programmes shows that this is not the case. No empirical study anywhere in the USA has ever found that government schools had worse outcomes when exposed to school choice. And there is a strong body of empirical evidence showing that school choice makes government schools better, not worse. The fears that public schools would be harmed by school choice simply have failed to materialise.

The research consistently has found that, where students can use school choice to attend any school, government or privately owned, the government schools make bigger academic improvements. Four studies of a school choice programme in Florida have found that government schools eligible for vouchers made dramatic improvements relative to other government schools. Three studies of Milwaukee's voucher programme found that government schools whose students were eligible for school choice made larger academic gains than other government schools. Studies of school choice programmes in Maine, Vermont and Texas confirm these findings (see Forster, 2007).

These studies are not able to use random assignment methods. Instead, they track individual-level or school-level data over time. Although this is not quite the gold standard, we can still have confidence in the results of these studies.

The finding that school choice improves government schools is counterintuitive to many people, and they have a hard time believing what the empirical evidence clearly shows. One reason school choice might be expected to improve government schools is because it allows parents to find the right school for each individual child. Every child is unique and has unique educational needs, and no one school can be the right school for every child. Another reason is that school choice does not actually drain money from school budgets, as we will see in more detail below. Finally, school choice provides positive incentives for improvement that are lacking in the traditional monopoly system. When government schools know that students can leave, using school choice if they are not getting an education, those schools have a much more powerful incentive to improve their performance and keep those students from walking out of the door.

Effects on other types of outcomes

The public is rightly concerned about other educational outcomes besides academic attainment. Racial segregation in schools has long been a prominent concern in the USA and has growing relevance in other countries as well. Schools are also expected to produce students who have strong civic values and provide appropriate services to students with disabilities. The fiscal impact of school choice programmes is also a legitimate issue. Empirical evidence is available on how school choice intersects with each of these concerns.

Racial segregation

It often is claimed that school vouchers lead to greater racial segregation. However, this claim is not often checked against the available evidence. In fact, the evidence is all on the other side: voucher programmes provide a greatly reduced level of racial segregation by breaking down neighbourhood barriers.

Contrary to many people's intuitions, there are good reasons to expect that school vouchers will reduce segregation. In the USA, under the monopoly system, school attendance is determined by where people live, so government schools inevitably reproduce the segregation that arises from segregated housing patterns. Privately owned schools, by contrast, typically draw students from a larger geographical area.

As with the research on academic outcomes, there is a great deal of research on segregation that uses inappropriate empirical methods. I have found a total of seven studies that used valid empirical methods to compare segregation levels in government schools with segregation levels in privately owned schools participating in voucher programmes. All seven studies found that segregation levels were lower in voucher schools than in government schools (see Forster, 2006).

Civic values

Another common claim is that privately owned schools do not do as good a job as government schools of teaching students to have good civic values, such as tolerance for the rights of others. This claim, too, is not often checked against the available evidence – which, as with the claims discussed above, runs in the other direction.

Just as many people find it counterintuitive that vouchers result in lower levels of segregation, many have difficulty believing that privately owned schools could do a better job of teaching tolerance and democratic values. However, there are several reasons this might be the case. One is that these schools are simply better at teaching, as the evidence discussed above shows. Another possibility is that privately owned schools, which often grow organically out of cultural traditions, can provide students with cultural roots; a considerable body of research has found that individuals who are secure in their own cultural identities are more likely to tolerate the different cultural identities of others. Privately owned schools also may benefit from being legally permitted to have a point of view on controversial subjects, rather than having courts constantly looking over their shoulders to make sure they remain 'neutral' (whatever the court decides that means) on all subjects of any controversy. This regime may breed a strong reluctance in government schools to allow controversial issues to be raised in the classroom at all – which would make it much harder for them to convey a tangible sense of what tolerance really is and why it is needed.

Wolf (2007) recently published a literature review that identified 59 findings from studies that compare civic values in government and privately owned schools. Of these, 23 findings used random assignment (taking advantage of random lotteries to admit applicants to voucher programmes) or other

highly rigorous methods that removed most of the influence of student selection into private schools. The other 36 used more basic methods. Of the 23 especially rigorous findings, 11 found better civic values in privately owned schools, 11 were neutral and only one found better civic values in government schools. The 36 more basic findings broke down into 20 finding better civic values in privately owned schools, 13 neutral and two finding better values in government schools.

The most commonly studied question on civic values was whether students show tolerance for the rights of others. Such studies typically ask students to identify their least-liked group. Students often pick groups such as the Ku Klux Klan, Nazis, Communists, pro-life or pro-choice groups, gay activists or the religious right. Students then are asked whether they would be willing to let members of this least-liked group engage in political activities such as marching in their town, running for elected office or having a book sympathetic to its views in the local library.

Wolf identified 13 highly rigorous findings on tolerance, of which eight were neutral and five found higher levels of tolerance in privately owned schools. He also identified eight more basic analyses of tolerance, of which six found more tolerance in privately owned schools, one was neutral and one found more tolerance in government schools.

Disability services

Services for students with disabilities are another common area of concern for school choice. The government school system maintains a large and costly bureaucracy whose purpose is to deliver special-education services. Since privately owned schools do not have a similarly large and visible special-education bureaucracy, many people assume they do not provide special-education services.

Studying outcomes for students with disabilities is even more difficult than studying most educational subjects. Since student disabilities run the gamut from mild to severe, and the exact severity of each student's disability is difficult to quantify, it is hard to measure how well schools are doing relative to how well they could be doing, given the students they have. Furthermore, there is not even a consensus on what measurements are appropriate for evaluating the academic achievement of students with disabilities.

However, at least one study has compared special-education services in government and privately owned schools. Florida's McKay programme allows any disabled student in government schools to move to a privately owned school using a voucher. An empirical evaluation of the programme I conducted with Jay Greene of the University of Arkansas (2003) compared the services these students had received in their previous government schools with the services they received in privately owned schools through the voucher programme. Parents reported much higher rates of satisfaction with their children's academic progress and services received in privately owned schools; students also were victimised by their peers less often and less likely to exhibit behaviour problems. Students were served about the same regardless of race, income or disability type.

To ensure that students who had unsatisfactory experiences would be included, we also collected data on the roughly 10% of families that had been in the programme in the previous year but were no longer participating. These former participants also reported that their privately owned schools had served them better than their previous government schools. More than 90% of them said the programme should continue for others, even though they were no longer using it themselves.

Fiscal effects

Finally, one of the most frequent complaints about school choice is that it drains money from government schools. This seems plausible on the surface – some amount of money from the state treasury (or from tax receipts, in the case of tax-credit scholarship programmes) that would otherwise have gone to government schools is going to support students in privately owned schools instead. However, the actual fiscal effect of school choice on government schools, and on state budgets, is a more complicated story.

US schools are funded by a combination of federal, state and local revenue, and federal revenue makes up only a small portion of the total. In a typical school choice programme, state funds associated with participating students are redirected, but local funds remain in the local school districts even after students have left. This is because local school funding is not closely tied to enrolment and doesn't change when enrolment changes. Government schools therefore lose only part of the funding that goes with each school choice student. But they lose all of the student, and therefore all of the student's costs. In other words, school choice reduces government schools' costs more than it reduces their revenues – saving them money.

Critics of school choice often counter that schools have fixed costs that don't go down when students leave – keeping the lights on in the school building and so forth. This is certainly true, but the savings produced by school choice are typically much larger than any plausible estimate of fixed costs.

School choice also saves money for state budgets. The amount of money a state spends per student in a school choice programme is typically less than the state portion of government school spending. For example, if the state portion of public school spending is $6,000 per student and the state offers students a $5,000 voucher, every voucher student saves the state $1,000.

A national study by Susan Aud of the Friedman Foundation (2007) has examined the fiscal effects of every existing school choice programme, going back to the founding of the Milwaukee voucher programme in 1990. To ensure a generous allowance for fixed costs, the study counts only savings in the variable category of 'instructional' expenditures, rather than in the total school budget. This is an overly conservative assumption, since many categories of spending other than instruction are known to be made up predominantly of variable costs rather than fixed costs.

The study found that, from 1990 to 2006, school choice saved $422 million for local school districts. It also saved $22 million for state budgets. This finding has been confirmed by other fiscal analyses of proposed school choice programmes in numerous states.

Evidence matters

The evidence on school choice does not answer all questions. It is subject to some methodological limitations. On some issues, we don't have as much evidence as we would like. And the benefits of school choice identified by these studies are sometimes moderate in size – not surprising, given that existing school choice programmes are restricted to small numbers of students and limited to disadvantaged populations, hindering their ability to create a true marketplace that would produce dramatic innovation.

However, these caveats should not be permitted to obscure the strength and depth of the evidence supporting school choice. A large body of top-quality studies consistently shows that school choice produces higher academic achievement for the students who have the opportunity to use it. On this issue, the evidence supporting school choice is as strong as the evidence on any social policy question whatsoever. The available evidence also supports school choice on other issues.

The research consensus on these issues ought to be acknowledged and allowed to affect the public debate over school choice. For all the faith that the public has in science – faith for which we scientists should be grateful – the public and its opinion leaders still have a long way to go in learning what the science really says about education. But the 'disconnect' between the claims made about school choice and what the empirical evidence shows about it cannot last forever. The mythology that keeps the monopolists going is a house of cards that eventually will fall.

References

Aud, S. (2007) *Education by the Numbers: The Fiscal Effect of School Choice Programs, 1990–2006*, Indianapolis, IN: Friedman Foundation for Educational Choice.

Barnard, J., C. Frangakis, J. Hill and D. Rubin (2003) 'Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City', *Journal of the American Statistical Association*, 98, 462, 299–306.

Forster, G. (2006) *Freedom from Racial Barriers: The Empirical Evidence on Vouchers and Segregation*, Indianapolis, IN: Friedman Foundation for Educational Choice.

Forster, G. (2007) *Monopoly versus Markets: The Empirical Evidence on Private Schools and School Choice*, Indianapolis, IN: Friedman Foundation for Educational Choice.

Greene, J. and G. Forster (2003) *Vouchers for Special Education Students: An Evaluation of Florida's McKay Scholarship Program*, New York: Manhattan Institute for Policy Research.

Howell, W. and P. Peterson (2002) *The Education Gap*, Washington, DC: Brookings Institution (revised edition 2006).

Howell, W. and P. Peterson (2004) 'Voucher Research Controversy', *Education Next*, 4, 2, 73–78.

Krueger, A. and P. Zhu (2004) 'Another Look at the New York City School Voucher Experiment', *American Behavioral Scientist*, 47, 5, 658–698.

Wolf, P. (2007) 'Civics Exam', *Education Next*, 7, 3, 66–72.

Wolf, P., B. Gutmann, M. Puma, L. Rizzo and N. Eissa (2007) *Evaluation of the D.C. Opportunity Scholarship Program: Impacts after One Year*, Washington, DC: US Department of Education.

Greg Forster is a senior fellow at the Friedman Foundation for Educational Choice. This article is adapted from his recent study, 'Monopoly versus Markets: The Empirical Evidence on Private Schools and School Choice' (greg@friedmanfoundation.org).