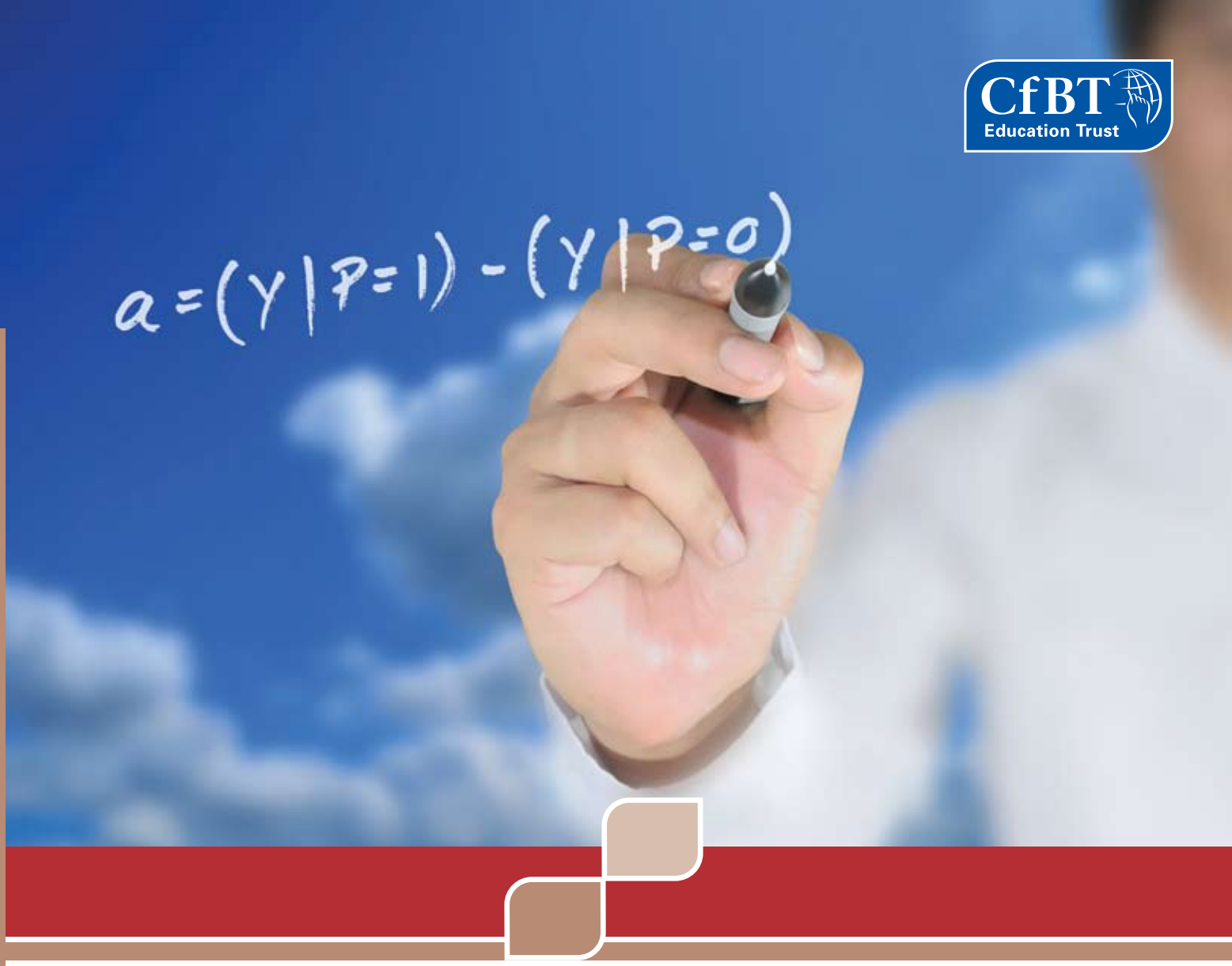


Impact evaluation of private sector participation in education

Research report

Laura Lewis
Harry Anthony Patrinos

Providing education for public benefit worldwide

Welcome to CfBT Education Trust

CfBT Education Trust is a top 50 charity providing education services for public benefit in the UK and internationally. Established over 40 years ago, CfBT Education Trust now has an annual turnover exceeding £100 million and employs 2,300 staff worldwide who support educational reform, teach, advise, research and train.

Since we were founded, we have worked in more than 40 countries around the world. Our work involves teacher and leadership training, curriculum design and school improvement services. The majority of staff provide services direct to learners: in nurseries, schools and academies; through projects for excluded pupils; in young offender institutions; and in advice and guidance centres for young people.

We have worked successfully to implement reform programmes for governments throughout the world. Government clients in the UK include the Department for Education (DfE), the Office for Standards in Education, Children's Services and Skills (Ofsted), and local authorities. Internationally, we work with education ministries in Dubai, Abu Dhabi and Brunei among many others.

Surpluses generated by our operations are reinvested in educational research and development. Our research programme, Evidence for Education, aims to improve educational practice on the ground and widen access to research in the UK and overseas.

Visit www.cfbt.com for more information.

Welcome to the World Bank

The World Bank is a vital source of financial and technical assistance to developing countries around the world. Our mission is to fight poverty with passion and professionalism for lasting results, and to help people help themselves and their environment by providing resources, sharing knowledge, building capacity and forging partnerships in the public and private sectors.

Contents

Foreword

from Neil McIntosh, CfBT Education Trust

The global outlook has moved from Education for All to Learning for All. In order to improve educational outcomes, governments are looking for innovations from both within and beyond the state sector. The non-state sector includes a diverse number of providers: community groups, non-governmental organisations (NGOs), faith-based organisations, trade unions, private companies, small-scale informal providers and individual practitioners. These suppliers may offer alternative school provision that meets the needs of individual students. Although the evidence is limited, there is some impact evidence that non-state providers particularly meet the needs of disadvantaged groups.

This paper outlines the need for greater investment into impact evaluation, particularly to evaluate these state-funded non-state providers. It aims to raise the awareness of impact evaluation amongst policymakers and other key stakeholders, especially parents. This will allow evidence-based policymaking and enable parents to make evidence-based decisions when choosing schools.

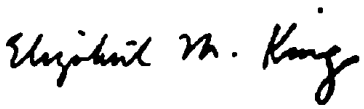
Neil McIntosh
Chief Executive, CfBT Education Trust

Foreword

from Elizabeth King, The World Bank

This paper is hugely important because it recognises the role that the private sector can play, not only as a provider of auxiliary services in school systems, such as transportation for students or school meals, but also as a provider of core learning services alongside state schools. It reminds us that although most learning institutions are state provided or financed, especially at the basic level, nearly all education systems include a range of formal and non-formal learning opportunities available to children, young people and adults in a given country, many of which are provided by the private sector. The paper is also important because it illustrates how impact evaluation methods can be applied to private provision in order to provide the kind of rigorous evidence that could be useful for programme design and policy choices, and it summarises the findings of key examples of impact evaluation.

Compared with two decades ago, more young people are entering school, completing the primary education level, and pursuing secondary education. Thanks to a combination of effective policies and national investments in education, far fewer children in developing countries are out of school. But this progress is not only the product of government efforts. Private-sector provision of education has been sparked by the limited ability of governments to meet the growing demand for education in some areas. Impact evaluation can help improve our understanding of how these investments can be more effective in addressing the education needs of countries.



Elizabeth King
Director, Education
The World Bank

About the authors

Laura Lewis is Lead Advisor, Education Performance at CfBT Education Trust. She is a leading expert on global education systems, with a particular interest in benchmarking and system transformation. Laura has worked with the Ministry of Education in Oman, the International Finance Corporation and the Asian Development Bank. She is currently working with the World Bank on its 'System Assessment Benchmarking for Results' programme (SABER) as well as an online toolkit for private-sector engagement with education. The toolkit has been developed to support policymakers in exploring the potential role for non-government organisations in improving student outcomes. Laura joined CfBT from McKinsey and Company.

For more information on the toolkit, see **http://www.cfbt.com/epsetoolkit/**

Harry Anthony Patrinos is Lead Education Economist at the World Bank, where he leads the Economics of Education Thematic Group and the East Asia System Assessment and Benchmarking Education for Results initiative. He focuses on labour market outcomes, school autonomy and accountability, demand-side financing and public-private partnerships. He manages several impact evaluations, including the school-based management experiments in Mexico. He has published more than 50 journal articles and co-authored several books, including *Making Schools Work* (with Barbara Bruns and Deon Filmer), *The Role and Impact of Public-Private Partnerships in Education* (2009), *Decentralized Decision-Making in Schools* (2009) and *Indigenous Peoples, Poverty and Human Development in Latin America* (2006). Mr Patrinos previously worked as an economist at the Economic Council of Canada. He has a PhD from the University of Sussex.

1 Introduction

> " 'Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink. Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou seest, deal with thy servants.' So he consented to them in this matter, and proved them ten days. And at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king's meat. Thus Melzar took away the portion of their meat, and the wine that they should drink; and gave them pulse." (From Chapter 1, Book of Daniel, King James Bible)

In many countries worldwide, there is public disquiet about the outcomes of school education. In response, many innovative approaches are being used to improve the quality of education. One such approach involves non-state organisations in the setting-up and operation of state-funded schools. Typically, such schools have more decision-making power than conventional state schools. Such autonomous schools are expected to have a positive impact on academic performance and the numbers of pupils staying on in education. But since such programmes are likely to be controversial and receive more scrutiny than other education programmes, it is therefore imperative that they be subject to rigorous impact evaluation. The paper gives a high-level overview of how impact evaluation can be used in the particular context of government-funded privately-provided schools. It also illustrates how more generally impact evaluation can support evidence-based policymaking. The paper is not a definitive guide and we provide references to further in-depth reading.

Around the world, policymakers are looking for innovative ways to improve school quality. International tests such as the OECD's PISA or IEA's TIMSS have allowed governments not only to look inwards at performance, but also to compare the quality of education with that in other countries. A recent report by the OECD (2010) found that raising performance by 25 points over the next 20 years would lead to an aggregate gain in GDP of $115 trillion over the lifetime of the generation born in 2010.

The private sector could support the Government in raising the quality of education. The private sector in its broadest sense includes communities, non-governmental organisations (NGOs), faith-based organisations, trade unions, private companies, small-scale informal providers and individual practitioners; all may collaborate with Government in order to raise education quality. Though the public sector remains the dominant player in education, making high quality education accessible for all in a country requires innovative programmes and initiatives in addition to public resources and leadership – or 'vision'. There are ways in which the public and private sectors can join together to complement each other's strengths in providing education services, helping countries to meet their education goals and to improve learning outcomes (Patrinos et al. 2009). These partnerships can be tailored and targeted to meet the needs of specific communities.

Private provision of public education services can produce several real benefits. Such benefits may include:

- competition in the market for education
- autonomy in school management
- improved standards through contracts
- risk-sharing between government and providers (Patrinos et al. 2009).

To ensure the benefits of private provision the Government must ensure that there is an effective regulatory environment. The regulatory environment must support strong accountability, informed parents, autonomous schools and competition. In a strong accountability system, autonomy is accompanied by responsibility. Policymakers hold all schools to account for the quality of their education and the accountability mechanism is transparent. Informed parents use both their voice and if the system permits, choice, to ensure the schooling supplied to their children is of the highest quality. These parents also hold the Government accountable through the political process whether it is local, regional or national.

Strong accountability and informed parents ensure that equity is protected. Some have argued that under a choice system private providers will be unaccountable to taxpayers and the public. Claims of efficiency gains are also questioned. Moreover, choice may lead to privatisation, less public (government) control of education, and increased segregation (Ladd 2002). Therefore, it is important that education systems are regulated adequately, providing information to parents and ensuring that there is a rigorous accountability process in place.

The private sector can offer alternative autonomous schools and enhance the competitive environment:

- *autonomous schools* that can tailor teaching and learning to meet the needs of all students they serve. Schools have control over the quality of educational professionals in the school.
- *a competitive environment* where schools offer a range of models to meet the needs of individual students.

Engagement of the private sector is often promoted as an innovative means of improving school quality. It is argued that competition for students will lead to efficiency gains, as schools – state and private – compete for students and try to improve quality while reducing cost (Friedman 1955; Hoxby 2003a; Neal 2002). The idea is that, when private schools are encouraged to attract students who otherwise would be educated in state schools, they become innovative and thereby bring improvement to the learning process. Likewise, state schools, to attract students and the resources that come with them, seek to improve and provide an education at least as good as the private schools. Thus, it is argued that school choice will lead to improved learning outcomes and increased efficiency in both types of schools. The OECD PISA has found school autonomy to be a considerable school-level factor associated with improved educational performance.

The Government has several options to engage the private sector. Public finance of education remains the Government's responsibility, but the provision of schooling need not be public. In fact, there are several ways that Government can ensure schooling, but not provide it. Publicly-financed, privately-provided education is one such option. Private management of public institutions is another. In addition, there are several mechanisms of private finance involvement in the education system. Figure 1 shows the options available; this paper is focused on the third quadrant (public finance, private provision).

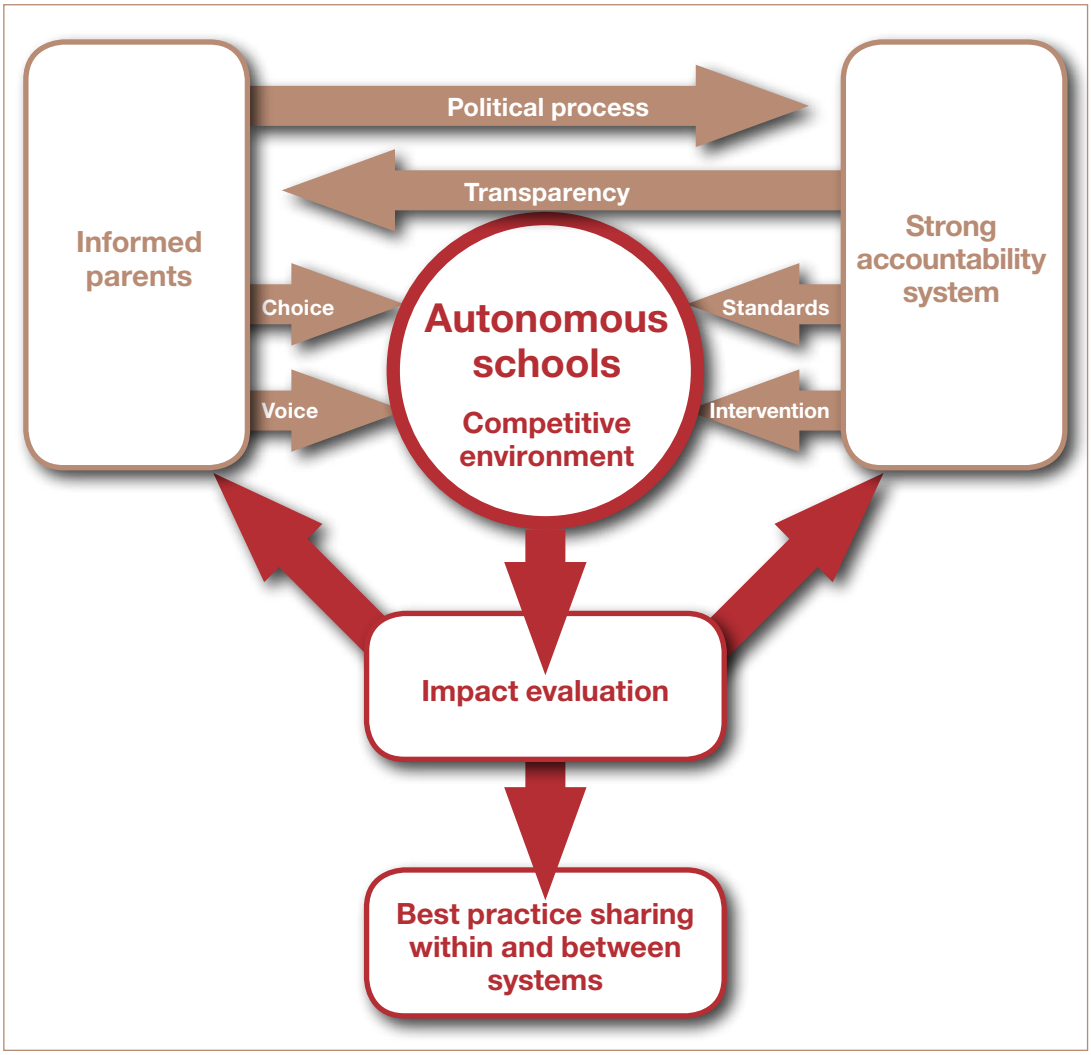
Figure 1: Provision and finance of education

		Provision	
		Private	**Public**
Finance	**Private**	I. Independent private school Private university Home schooling Tutoring	II. User fees Student loans
	Public	III. Private funded Private contracted Private management (Charters) Market contracted (Vouchers)	IV. Traditional public (state) school Public university

Source: Adapted from Patrinos et al. 2009; for full definitions of private sector engagement in the third quadrant, see Annex 1, page 33.

There is no substitute for a rigorous impact evaluation. These claims and counter-claims are empirical questions that need to be settled through a rigorous analysis of the evidence. Impact evaluation can focus on both the educational outcomes of the privately-provided schools and may also look at the effect on other schools within the competitive environment. Figure 2 shows how impact evaluation can support an effective regulatory environment, inform evidence-based policymaking and ensure that best practice is shared between systems.

Figure 2: Effective regulation and the role of impact evaluation



Source: Adapted from Lewis and Patrinos (2011)

There is a need to increase the number of impact evaluations to create a rigorous evidence base in order to inform policy. An evidence base is required which evaluates impact on a case-by-case and country-by-country basis. Evidence on the impact of interventions that involve the private sector in the provision of education services is growing, as is the evidence of education interventions in general. Most of the rigorous impact evaluations are coming from the United States. There are a few evaluations from other countries and even fewer randomised trials; but the evidence base is growing (Barrera-Osorio and Patrinos 2009; Patrinos et al. 2009).

2 The importance of generating evidence of impact[1]

James Lind, the Scottish physician, conducted the first-ever clinical trial and developed the theory that citrus fruits cured scurvy. While it is now known that scurvy is a disease caused by a deficiency of Vitamin C, in the eighteenth century it caused many deaths. Lind was not the first to suggest citrus fruits as a cure for scurvy, but he was the first to study their effect by a systematic experiment in 1747. On a ship that was afflicted with scurvy he divided twelve sailors into six groups. He gave each group different additions to their basic diet. Some were given cider, others seawater, others a mixture of garlic, mustard and horseradish. Another group of two were given spoonfuls of vinegar, and the last two oranges and lemons. Those fed citrus fruits experienced a remarkable recovery. By 1800 ships began to regularly carry lemon juice (Hughes 1975).

Impact evaluation identifies the effect of an intervention on outcomes. Impact evaluation techniques compare the impact on the beneficiaries of a certain policy intervention or project with a counterfactual group that has not been exposed to the same intervention or project. The results from impact evaluations can help inform policymakers on where to allocate scarce resources and can provide evidence on whether current policies are working or not (Gertler et al. 2010).

Impact evaluation seeks to estimate the effect of a given intervention on a critical outcome (such as enrolment or performance in standardised tests). Observing the same individual with and without the programme creates a fundamental problem, given that we will never observe the same individual in two different states at the same time. Therefore, impact evaluation tries to assess the impact of programmes through the application of a range of analytical techniques. The key techniques are described in section 3.

In an age of scarce resources, public policy requires evidence of 'what works'. A rigorous impact evaluation can show if things work and why they work. Effective evaluation will inform policymakers and permit improvements in the implementation of policies and programmes. Evaluation can inform programme design (for example, eligibility and types of benefits) and can improve operations and efficiency. Additionally, the information generated by the impact evaluation may be useful for programme sustainability and can be a valuable asset in the negotiation of budgets and in the provision of reliable information to inform public opinion. Without reliable information on the effects of education initiatives there can be no accountability or understanding of return on investment.

In many countries private participation in school education is highly contested and politically controversial. It is a topic that generates considerable debate and there is a need for evidence to inform the debate. Given this, it qualifies as an important area worthy of rigorous impact evaluation. More importantly there is a dearth of rigorous evidence on private participation, and globally important policy questions are left unanswered.

[1] This section draws heavily on the published literature, especially Gertler et al. (2010), Gertler, Patrinos and Rubio-Codina (2007), and Barrera-Osorio and Patrinos (2009), to which the reader is referred for more details.

3 An overview of impact evaluation techniques

Impact evaluation attempts to answer certain key questions: What was the effect of the programme on outcomes? How much better off are the beneficiaries because of the programme/policy? Typically, one would ask: How much does an education programme improve test scores (learning)? What is the beneficiary's test score with programme, compared to without programme?

It is possible to represent the effect size of an intervention through an equation:

$$a = (Y \mid P=1) - (Y \mid P=0) \qquad (1)$$

where the net effect is presented as a, Y represents outcomes, and P is the programme or intervention (Gertler et al. 2010), where P=1 are students who entered the programme and P=0 are students who did not enter the programme. In an ideal world the researcher would compare the same individual with and without the programme at the same point in time. This of course is impossible since we cannot clone a person. This is the evaluation problem, and solving it requires that we come up with the valid 'counterfactual': what would have happened without the programme? The concept of the counterfactual is important and deserves further clarification. An observed treatment is given to a person and the outcome of that treatment is Y(1). The counterfactual is the outcome that would have happened, Y(0), if the person had not received the treatment. An effect is the difference between what did happen and what would have happened, or:

$$Effect = Y(1) - Y(0)$$

The estimated impact is the difference between the treated observation and the counterfactual. Since we cannot observe the same individual both with and without the programme at the same point in time, then we need to estimate the counterfactual. The counterfactual is the key to impact evaluation.

The criteria for identifying subjects for the treatment and the counterfactual analysis are that they have identical characteristics, except for benefiting from the intervention. There should be no other reasons for differences in outcomes of treated and counterfactual. The only reason for the difference in outcomes is due to the intervention.

There are several characteristic examples of invalid counterfactuals. These include a study of the same subject before and after an intervention; those not enrolled in school compared to those who are; and those not offered the programme compared to those who were offered the programme. The problem with this is that we do not know why they are not enrolled in the programme. Thus, there is the potential for selection bias: that is, people choose to participate for specific reasons, for example motivation. Many times those reasons are related to the outcome of interest. One cannot separately identify the impact of the programme from these other factors or reasons.

Creating a valid counterfactual requires attention to a range of important design issues. Another criterion for a valid counterfactual is that one knows all the reasons why someone receives the programme and others do not. Also, one needs to know all the reasons why individuals are in the treatment. However, if the reasons are correlated with the outcome then one cannot identify or separate programme impact from other explanations of differences in outcomes. One needs to be able to guarantee the comparability of treatment and counterfactual – also known as control – groups. With a proper counterfactual, then the only remaining difference is the intervention.

Randomisation

The best mechanism of guaranteeing a proper counterfactual and unbiased evaluation is *randomisation*. This method gives all an equal chance of being in the control or treatment groups. It guarantees that all factors and characteristics will be on average equal between the two groups. The only difference is the intervention.

Besides randomised assignment, or randomised control trial (RCT), other experimental methods include:

- *Randomised promotion*, or encouragement design, can be created by targeting a specific sub-population to receive the treatment. This creates an intent-to-treat instrument, based on the initial treatment intent, not on the treatment eventually administered; this preserves internal validity and yields an unbiased estimate about the effects of being assigned to treatment, not of receiving treatment; this may be of policy interest, but should be complemented by other analyses.
- *Regression discontinuity design* – which allows one to create treatment and control groups around an arbitrary cut-off point.

Quasi-experimental

While there are quasi-experimental designs, they require more data and it is more difficult to accurately ensure the methodologies employed produce unbiased estimates. The different approaches vary in feasibility, cost and the degree of clarity and validity of results. The design also determines the set of estimation methods available to obtain an unbiased estimate of the programme impact. The methodology employed will further depend on the type of information available, the underlying model, and the parameter of interest. For example, datasets with longitudinal or repeated cross-section information will support less restrictive estimators due to the relative richness of information. The main quasi-experimental methods include:

- *Matching, or propensity score matching* – which uses observable characteristics to create a group similar to the treatment
- *Difference-in-difference* – compares the pre- and post-intervention outcomes within each group and compares the within-group variation between the two groups. This requires the two groups to be judged equal in terms of characteristics and one has a way of attributing programme access to criteria unrelated to programme outcome
- *Programme phase-in* – a technique that exploits the phasing-in of a programme to create a treatment and control group
- *Instrumental variables* – a method of using proximate variables that determine choice but not outcome
- *Heckman selection models* – a two-stage method which allows the researcher to correct for selection bias (Heckman 1979)
- *Fixed effects* – a regression technique that controls for unobserved and time-invariant characteristics that may influence the outcome variable.

Quasi-experimental methods require researchers to identify plausible counterfactuals, which is more difficult than randomisation. While quasi-experimental methods can allow one to compare the results in the intervention group to a plausible counterfactual, the truth is that the farther one is from an experimental design, then the more un-testable assumptions one must adopt to use the method. Nevertheless, since failure to evaluate is not acceptable, then one must proceed with quasi-experimental evaluation while awaiting the opportunity for an RCT design, but keep in mind the caveats and adopt robustness tests in the analysis. A summary of the evaluation techniques is presented in Figure 3.

Figure 3: Summary of impact evaluation techniques

Experimental
Randomised control trial Randomised phase-in Encouragement
Regression discontinuity

Quasi-experimental
Difference-in-difference
Matching
Phasing
Instrumental variables
Fixed effects
Heckman selection

Small-scale evaluation tools

There are programmes that are just beginning or do not yet have sufficient scale or built-in mechanisms to make a national randomised control trial possible right away. For those instances, other rigorous evaluation techniques could be applied, though not all will allow causal inference of the effects of the programme. They may be proof of concept, or internal evaluations:

- *Small-scale randomised trial.* There is rigorous evidence that has been gained from small trials in one location. For example, the early childhood intervention made famous by the randomised trial involved only 123 children in the High Scope Perry Preschool programme in Ypsilanti, Michigan (Schweinhart, Barnes Weikart 1993). Another example is the recently published randomised evaluation of 457 students at a KIPP charter school in Lynn, Massachusetts (Angrist et al. 2010).
- *Effectiveness studies.* There are treatments that can work, but for a variety of reasons, they do not: adverse reactions, demanding schedules, denying treatment, and so on. This is the realm of effectiveness studies; that is, does the intervention work under real-life conditions?
- *Process evaluations.* These are at aimed at enhancing the programme by understanding it more fully. Process evaluations measure what is done by the programme, and for whom these services are provided. Ideally, process evaluations assist in the identification of 'active ingredients' of treatment, and assess whether a programme is meeting accepted standards. In general, process evaluations pose questions in two areas: coverage and process.
- *Within-group intervention.* This approach compares sub-groups within the intervention. It can focus on outcomes but could also be used to look at processes. For instance, schools could give information to some parents. The school could then look at the impact of information on parental engagement, student behaviours, student outcomes, and so on.

Non-random			Random
Effectiveness	Process	Within intervention	Small-scale randomised control trials

4 Applying rigorous impact evaluation to programmes that engage the private sector in the delivery of education

The main objective of an evaluation of any school intervention is to investigate the impact on educational outcomes. Usually, the educational outcomes are captured by variables such as drop-out, attendance, repetition and failure rates, as well as learning outcomes approximated by standardised tests. There are two main channels through which an engagement with the private sector in the delivery of education – let us call it choice, for shorthand – can increase learning outcomes:

1. The mechanism it uses to sort students
2. The across-school competition it creates (Barrera-Osorio and Patrinos 2009).

Of the first channel, advocates claim that choice programmes allow a better alignment of preferences between parents and schools. Parents, when given the option, will enrol their children in high-performing schools, either leaving or not applying to the low-performing ones. This allocation mechanism will support students to improve their educational outcomes because they will be in better schools and enjoy better peers.

Of the second channel, advocates argue that choice will induce competition for students across schools. Low-performing schools will be forced to increase their quality in order to retain and attract students who will otherwise enrol in better schools (Hoxby 2003a; Neal 2002).

The average learning outcome, then, will depend on the average outcomes of students who stay in the low-performing schools (expellers) and those who move to the high-performing schools (receivers). In a programme that allocates students randomly, the students leaving the expeller schools will have the same characteristics as the ones staying and will thus observe an improvement in their education. However, the receiving schools may observe a change in the average quality of education, depending on the new composition of students. That is, if the students entering the school perform at lower levels on average than the current students, the school's quality of education may decrease (Nechyba 1999, 2000; Epple and Romano 1998).

In programmes that do not allocate benefits randomly, the type of children leaving the low-performing schools are presumably the more able students, and the students remaining in the school are, therefore, the less able ones. Thus, the high-performing schools will receive the best students and the low-performing ones will keep the less able students. In other words, the average quality of the expeller schools will fall, and the average quality of recipient schools may rise or fall (Nechyba 1999, 2000).

On top of this sorting mechanism of students, competition across schools will tend to increase the quality of education. Again, schools will compete for students, which may lead to improved learning outcomes. Final outcomes will depend on the net effect of these margins. Building on equation (1), the typical estimation of the effect of a choice programme in learning outcomes will have the form:

$$Y_{i,j,t} = \beta_0 + \beta_1 * X_{i,t} + \beta_2 * Z_{j,t} + \beta_3 * P_{i,j,t} + \varepsilon i,j,t \qquad (2)$$

where $Y_{i,j,t}$ is any variable capturing schooling outcomes such as drop-out, repetition or standardised tests for individual i in school j at time t; $X_{i,t}$ are a set of socioeconomic characteristics of the individual; $Z_{j,t}$ are characteristics of the school such as infrastructure capacity or quality of teachers; $P_{i,j,t}$ is a dummy variable indicating whether the individual is a beneficiary of the programme or not; $\varepsilon i,j,t$ comprises all unobservable characteristics of the school and individual that can affect learning outcomes.

The fundamental problem in estimating the impact of choice programmes with equation (2) is *selection bias* – students and schools self-select into the programme. This phenomenon is also known as sorting bias. Under these conditions, a comparison between students who participate and those who do not confounds the effects of the programme, with the initial differences in characteristics between participants and non-participants. For example, it is possible to expect that better informed households are more likely to apply to choice programmes. In this case, students from these households may perform differently from students who did not apply to the programme. Therefore, any observed final educational outcomes not only comprise the results from the choice programme but also the inherent differences in characteristics of the families or students (Gertler, Patrinos and Rubio-Codina 2007).

Sorting bias can also affect treatment estimates on school test scores or other outcomes. If choice schools do a better job at retaining students who would otherwise have dropped out, then average school achievement remains lower. That is, the achievement effect is washed out by an attainment effect and underestimated. The converse is also possible if choice schools attract better-performing students. Controlling for this form of bias can be done using data on school rolls on enrolment, passing rates and desertion rates.

Besides students self-selecting into the programme, schools may also self-select into the programme, reinforcing the problem of identifying impacts. It is possible that some private schools may not want to accept programme students. An evaluation of the programme with this behaviour may thus confound the impact of the programme with the differences in the characteristics of the schools. Presumably, the schools that choose to enrol programme students are different from the other schools. Therefore, a simple comparison between students in schools with the programme and without the programme may pick up not only the differences in the educational outcomes due to the choice programme, but also the differences in the characteristics of the two groups of schools.

The evaluation question then, is to identify β_3. Or, put another way, how to construct a counterfactual with no intervention against which to measure the change with intervention. If the programme does not assign beneficiaries with a random mechanism and individuals and/or schools self-select into the programme, then it is very likely that the unobservable characteristics are correlated with the variable that indicates whether or not the individual receives the programme. In this case, programme estimates from typical techniques are biased.

This can be described as a problem of causality. If *P* is comprised of self-selected individuals, and the most able ones are participating in the programme, then *P* can be causing *Y*. On the other hand, *Y* as indicative of ability can be *causing P*, in the sense that more able individuals are choosing to be in the programme. Therefore, the causality direction is unclear.

Another perspective of the problem is to see bias as the consequence of omitted variables. In this case $\varepsilon_{i,j,t}$ captures all the unobservable variables at the school and individual level that affect *Y*. If it were possible to control for all the variables, then one would be able to produce unbiased estimates. In other words, if one could control for all key variables that determine participation in the programme, then estimates from typical procedures, such as regression analysis – any technique for modelling and analysing several variables when the focus is on the relationship between a dependent variable and one or more independent variables – would be valid and unbiased. However, since participation in a programme is also determined by unobservable characteristics such as the drive of the school principal, his ability to raise funds, and parental or governmental preferences, then typical estimates will suffer from *omitted variable bias*.

The direction of the bias in simple comparisons between students with choice and other students is not clear. For example, schools may decide to self-select into the programme, as is the case in Chile's nationwide choice programme that began in 1981. Elite private schools opted not to receive voucher students. Under the assumption that schools that do not participate in the programme are better than the ones that do, it is likely that a simple comparison between students in the voucher programme versus students without the voucher will tend to be downward biased. On the other hand, presumably students from better informed, active and motivated families are the ones that take advantage of the programme. In this case, the bias tends to be upward. The net effect – for example, the effect of school choice and student choice – is unknown.

Governments may want to target interventions to areas with particular needs and characteristics that are thus systematically different to those areas where the programme is not allocated. For instance, the Government may assign benefits to more disadvantaged schools first, given budget constraints. This would produce a negative correlation between the school unobserved components (captured in the error term in equation (2)) and the treatment variable. Hence, estimates of programme impact would be downward biased. On the other hand, governments may be just as likely to place treatment in areas that already have good education outcomes in order to increase the chances of positive outcomes or because they might derive political support from some groups. Alternatively, better-performing schools that have stronger and more concerned parents might push authorities to allocate benefits to their school. In either situation, these schools are likely to continue to do better than worse-performing and less influential schools even without the programme. Hence, programme impact estimates will likely be upward biased. Biases coming from this source are known as *endogenous programme placement bias*.

The problem of bias in the estimation of equation (2) has multiple solutions. Ideally, the voucher programme will be allocated using a lottery or a randomisation mechanism (Shadish, Cook and Campbell 2002; Duflo, Glennerster and Kremer 2007). If randomisation is not used, then regression discontinuity analysis can be used, given unbiased estimators (Hahn, Todd and Van der Klaauw 2001). The other potential methods – instrumental variables, Heckman correction models, difference-in-difference estimators and matching estimators – are based on strong assumptions (Angrist and Imbens 1995; Athey and Imbens 2006; Heckman 1976; Heckman, Ichimura and Todd 1998; Rosenbaum and Rubin 1983).

Experimental

Randomisation and regression discontinuity designs (RDD) produce unbiased estimators of programme impacts. Impact evaluation using randomisation strategies is based on the idea that a lottery will *de facto* create similar treatment and control groups in terms of observable and unobservable characteristics. In this sense, the mean of observable variables and unobservable variables will be equal across groups. The only difference between treatment and control groups is the intervention. Therefore, differences in outcomes can be attributed solely to the programme. In short, randomisation assures that estimated impacts are unbiased.

Randomisation

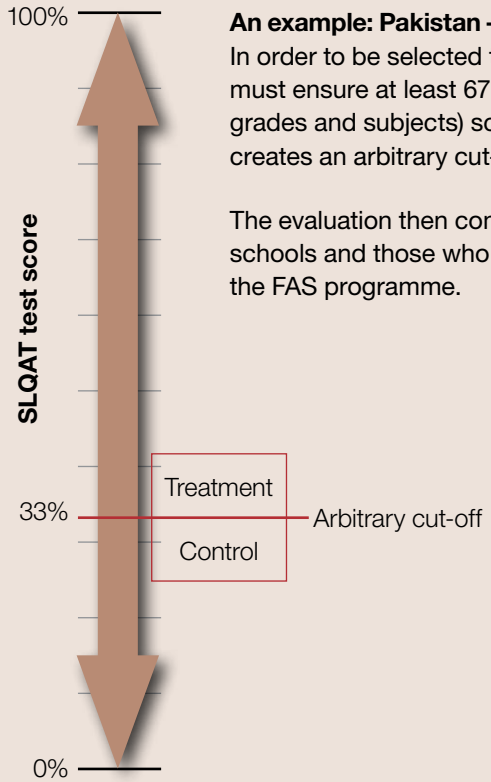
Randomised designs involve two stages: first, the identification of a group of potential beneficiary schools (or willing participant schools, or students) with similar characteristics; and second, the random allocation of schools (or students) to the treatment group or control group. This makes treatment and control schools virtually identical and causal inference feasible. Hence, a well-executed randomised experiment will greatly simplify the statistical analysis: the analysis will compare the mean difference in outcomes for the treatment and control groups to obtain consistent estimates of impact. For this reason, when using experimental data, the researcher should always assess the balance of the treatment and control groups, for example comparing the means in each group and/or the distribution of student, school and locality characteristics. Similarly, one should compare pre-intervention means or distributions of the outcome variables (this is known as the exogeneity test). The key advantage of randomised designs is that they overcome many of the problems encountered when using other evaluation practices without having to resort to difficult-to-test and hard-to-satisfy behavioural assumptions.

Lotteries have been used when there is over-subscription to randomly allocate either students or schools into the programme. When there are more quality applicant schools than available resources to allocate, or more students than can be adequately admitted, then it would seem fair to allocate benefits to particular winning schools and not to others using a lottery. (See Angrist and others (2002) on the evaluation of a school voucher programme in Colombia that benefited from a lottery design to facilitate the excess demand for school places.)

It could also be the case that the relevant authority temporarily lacks the capacity or resources to implement the intervention to all target schools. A natural solution would be to randomise the order in which the intervention expands to all potential beneficiaries. This design – known as ***randomised phase-in*** – would not only allow all schools to eventually benefit from the reform but would also provide the researcher with a clean source of identification. In a randomised phase-in, schools are selected but are assigned to the wave in which they will enter the programme. Entering the programme in different time periods allows for time series comparisons between those taking part in the intervention and the control group. The evaluation then compares student progress of those in the first wave with student progress of those who will enter the programme in the future. It is, however, important to allow enough time for the intervention to have effects amongst the initially treated before allocating benefits to the control group.

Regression discontinuity design

The regression discontinuity design (RDD) estimation is possible if the programme identifies its beneficiaries using a cut-off point. For instance, if the programme selects individuals based on a mean score (cut-off), then all those students with a score below a certain cut-off are beneficiaries of the programme and those with a score above the cut-off are denied access to the programme. Students with scores just below the cut-off point (beneficiaries) are similar to students above the cut-off point (comparison group). In this case, it is possible to compare the outcome variables for those two groups, and attribute the differences to the programme, given that one would expect the students in the two groups to be similar in their characteristics, with some just making it into the programme and the others just barely being denied access. Regression discontinuity analysis resembles a randomisation since from the point of view of the students, to be 'just below' or 'just above' the arbitrary cut-off point is almost like taking part in a lottery.



An example: Pakistan – Punjab Foundation Assisted Schools (FAS)
In order to be selected to take part in the FAS programme, schools must ensure at least 67% of tested students (pooled across the grades and subjects) score 33% or higher on the SLQAT test. This creates an arbitrary cut-off.

The evaluation then compares student outcomes between FAS schools and those who fell short of the criteria to be admitted onto the FAS programme.

Randomised promotion or encouragement design

Often experiments can be difficult to implement because it is difficult to ensure that all those selected for treatment receive the treatment and all selected for control do not. Also, it is sometimes impractical to conduct experiments. The encouragement design is a special case of an experiment that can be used in situations with little control over subjects' compliance. The main idea is that instead of randomising the application of the intervention itself, what is randomised is encouragement to receive the treatment. By randomising encouragement and carefully tracking outcomes for all those who do and do not receive the encouragement, it is possible to obtain reliable estimates of both the encouragement and the intervention itself (Diamond and Hainmueller 2007).

Encouragement may take the form of information that is additional to whatever is already part of programme implementation and targeted at student or parent level. Some subjects receiving encouragement may not follow through with the programme. Others who do not receive encouragement may nevertheless access the programme. All that is required is that the encouragement increases the likelihood that units will follow through with what they are being encouraged to do.

Encouragement involves randomly applying a certain type of intervention to a certain sub-group and not another, and comparing the differences in outcomes. For example, information or training is offered to a randomly encouraged group; the analysis then compares the improvement of outcomes for those randomly encouraged to those who were not encouraged. This is known as the *intention-to-treat* (ITT) effect and is the effect of the encouragement itself; for example, what is the effect of information on applying for a choice programme?

Since encouragement itself is randomised, the comparison between the encouraged (for example, better informed) and not-encouraged groups will be free of any bias due to self-selection if encouragement was delivered as planned. What is important here is that one can measure results for all subjects from the sample in which randomised selection was undertaken. Whenever encouragement takes the form of key project activities (for example, information campaigns) measuring the impact of encouragement is particularly useful.

One can estimate the effect of the treatment. This is done by exploiting a randomised encouragement design by adjusting the *ITT* effect by the amount of non-compliance. This yields to the *local average treatment effect*, or *LATE* (Imbens and Angrist 1994):

$$LATE = ITT / Compliance\ rate$$

where compliance rate = fraction of subjects that were treated in the treatment group minus fraction of subjects that were treated in the control group. If the compliance rate is 100 per cent, then LATE = ITT, we have perfect compliance, and all assigned to the treatment take the treatment and all those assigned to the control do not take the treatment. The compliance rate can be thought of as the fraction of subjects that fall into the sub-population of 'compliers', the group for whom the decision to take treatment was directly affected by the assignment. Put differently, this is the group induced by the encouragement to take advantage of the treatment.

An example: Assume the Government wants to target the underprivileged. It encourages this target group through increased information. This encouragement is randomly allocated within the target group. The researcher can then evaluate as follows:

- Comparing the student outcomes of those who are given information and act on the information, and the student outcomes of those who are in the same target group but did not receive the information (intent to treat).
- Reviewing the quality of the encouragement provided, by looking at the compliance rate, what percentage of those encouraged actually chose the intervention?

Quasi-experimental

Matching

The technique tries to capture the characteristics of intervention schools and matches them to schools with similar characteristics but not in the intervention. A widely used method of matching is *propensity score matching* (PSM). The propensity score is generally a function of observable school characteristics such as school type and size and locality characteristics. Instead of finding the best comparison school(s) for every single characteristic, one defines (the neighbourhood of) similar school(s) based on their propensity score. Each treated school is then paired with its selected set of non-participant schools. PSM techniques can also be applied to address self-selection biases at the student level.

There are two main challenges with the use of matching methods. The first is related to the heavy requirements they impose on the data. Exhaustive information on the characteristics of participant and non-participant schools is needed to model the participation decision. But the more detailed this information is, the harder it is to find a similar comparison group, as treatment and comparison schools will have to be matched on a larger number of similar characteristics. That is, there is a trade-off between the quantity of information to use and the size of the comparison group. A common approach is to construct the propensity score using pre-intervention characteristics. However, this requires the existence of exhaustive pre-intervention (baseline) data for both the participant and non-participant schools, which is often not the case.

Difference-in-difference

Difference-in-difference compares the before-and-after difference in the treatment group with the difference before and after of the control group. This approach requires pre- and post-intervention data (though not necessarily for the same schools). Its main advantage over matching methods is that individual characteristics that determine (participation into) treatment no longer bias impact estimates since they are differenced out in the estimation equation. Firstly, it requires access to long-term series of pre-intervention data from both types of schools to compare pre-trends over long enough periods. Secondly, it does not control for any time-varying school characteristics that affect (participation into) treatment. In applying this method, one must control for as many time-varying observable characteristics as are available and include separate time trends for treatment and comparison schools in the estimation in order to minimise the potential for biases.

Programme phase-in

Differences in the timing of programme implementation in different schools or in different geographical areas (school districts, localities, states) can also facilitate forming comparison groups. Examples of exogenous variations of this sort are administrative delays in programme implementation or the application of a time-varying geographic targeting rule uncorrelated with outcomes. In these situations, the implementation of the intervention automatically generates a valid counterfactual net of potential self-selection biases, as both participant and not-yet-participant schools are potential beneficiaries and are thus likely to have similar observed and unobserved characteristics. It is then possible to compare schools that are already being treated with schools that will be treated in the future using matching methods or simple differences.

Instrumental variables

This method is used to estimate causal relationships when controlled experiments are not feasible. This allows for consistent estimation when the explanatory variables are correlated with the error terms of a regression relationship. In cases where an instrument is available, consistent estimates may be obtained if the instrument is a variable that does not itself belong in the explanatory equation. However, plausible and valid instruments are difficult to find (Heckman 1979; Blundell and Costa Dias 2000; Davidson and MacKinnon 2003).

Summary

The best option for estimating the causal impact of a programme is a randomised control trial. This is not to say that everything we want to know can be ascertained this way; but in terms of first-generation questions – such as the blunt question, Does it work? – there are not many other options. Thus, randomisation is the preferred option.

5 Examples of school choice and private participation programmes using rigorous impact evaluation in the existing evidence base

Randomisation

The United States has been at the forefront of market reforms in education. Though several evaluations of choice policies and programmes have been undertaken, until recently they have seemingly produced mixed results. The Center for Research on Education Outcomes (CREDO 2009) used data on over 70 per cent of students in charter schools nationwide across 15 states and Washington, DC. When comparing these charter school students to their ***matched*** twins on mathematics and reading growth, the study found mildly negative effects of charter schools relative to traditional state schools. These are largely consistent with the findings of other studies of charter school performance which use student ***fixed effects*** [a statistical way of controlling for omitted variable bias, so-called because it holds constant *(fixes)* the average differences between the determinants of a variable by using dummy variables; thus, for example, the effect of school location on school costs might be controlled by having a dummy variable to represent each school's location and this would pick up the (often unobservable) effects of otherwise omitted variables (local wage effects, distance from markets, etc.)] The CREDO study's findings were criticised for potentially not adjusting enough for selection bias (Hoxby 2009).

However, some recent studies show positive results. These encouraging results come from studies using ***random*** lotteries commonly used to select which students can attend over-subscribed charter schools. A study from Boston demonstrated that charter schools raised mathematics scores by more than half a standard deviation per year in middle school (Abdulkadiroglu et al. 2011). Findings from Chicago and New York were more modest but still statistically and economically significant (Hoxby and Murarka 2009; Hoxby and Rockoff 2005). A study of the Harlem Children's Zone, which focuses on the poorest minority students, finds that students enrolled in grade 6 gain more than one standard deviation in mathematics, and up to half a standard deviation in English by grade 8 (Dobbie and Fryer 2011). The authors argue that these effects are enough to reverse the black-white achievement gap in mathematics. Another charter provider, the Knowledge is Power Programme (KIPP), which also focuses on low-income and minority students, produced gains of 0.1 standard deviations a year, with larger gains for special education and limited-English students, of the range 0.3 to 0.4 standard deviations a year (Angrist et al. 2010). These findings demonstrate the benefits of randomisation and they offer some clues on the mechanisms by which charters improve results – the 'no excuses' model.

Outside the United States, there are a few evaluations that take advantage of a randomisation design. An exception is the secondary school voucher programme (PACES) in Colombia. Due to over-subscription in the programme available places were allocated by lottery. This created a natural, randomised experiment that enabled researchers to undertake rigorous impact evaluations of the programme and test several hypotheses. The Colombian case, launched in 1991, provided the poorest third of its population with access to secondary education. The programme, which ran until 1997, covered 125,000 children. The vouchers were renewable through to the end of high school as long as the student continued to progress through school. More than three-quarters of the beneficiaries renewed their vouchers. The vouchers could be used at private academic and vocational schools and about 40 per cent of private schools accepted the voucher. The unit costs for participating private schools were 40 per cent lower than for non-participating private schools.

The results for this targeted voucher programme are encouraging. Researchers found that voucher beneficiaries had higher educational attainment: they were 10 per cent more likely to finish the 8th grade three years after they won the vouchers. They were also 5 to 6 per cent less likely to repeat a grade. They scored 0.2 standard deviations higher on achievement tests than non-voucher students. And they were 20 per cent more likely to take the college entrance exam than students who had not won a voucher in the lottery. They were also 0.6 to 1.0 per cent less likely to be married and 2.5 to 3.0 per cent less likely to be working (Angrist et al. 2002). In a study of longer-term effects, Angrist et al. (2006) found that the programme improved scores for both average students and those over the 90th percentile. Another study tested whether vouchers increase educational productivity or are purely redistributive, benefiting recipients by giving them access to more desirable peers at others' expense. Among voucher applicants to vocational schools, lottery winners were less likely to attend academic secondary schools and thus had peers with less desirable observable characteristics. Despite this, lottery winners had better educational outcomes. In this population, vouchers improved educational outcomes through channels beyond redistribution of desirable peers (Bettinger, Kremer and Saavedra 2010).

Korean elementary school graduates are randomly assigned to either a public or a private middle school in the residence district. This allows Kang (2007) to apply the randomisation to test whether students' learning outcomes are influenced by peer effects. In a strict sense, Kang (2007) does not study the effects of voucher programmes; instead, he tests peer effects using the randomisation of the programme. Kang (2007) finds that peer effects do indeed affect individual performance. This leads to the finding that a one standard deviation increase in the mean quality of one's peers enhances mathematics scores by a significant 0.4 of a standard deviation.

A programme designed to stimulate girls' schooling through the creation of private girls' schools in poor urban neighbourhoods of Quetta, Pakistan proved to be effective. Enrolment growth in these randomly selected neighbourhoods is compared to enrolment growth in otherwise similar neighbourhoods that were randomly assigned to a control group. The analysis indicates that the programme increased girls' enrolment around 33 percentage points. Boys' enrolment rose as well, partly because boys were allowed to attend the new schools and partly because parents would not send their girls to school without also educating their boys. This outcome suggests that programmes targeted at girls can also induce parents to invest more in their boys (Kim et al. 1999).

Regression discontinuity

When programme placement provides clear-cut rules on allocation, then researchers can use this information to create valid treatment and control groups. A few examples show the sort of interesting results that can be produced. School choice is an intervention often promoted as a means of improving efficiency. In Tel Aviv, Israel, a programme terminated an existing inter-district bussing integration programme and allowed students free choice among public schools (Lavy 2010). The researcher based his identification of the programme on ***difference-in-difference*** and ***regression discontinuity*** designs to yield various alternative comparison groups drawn from untreated tangent neighbourhoods and adjacent cities. Across identification methods and comparison groups, the results consistently suggest that choice significantly reduces the drop-out rate and increases the cognitive achievements of high-school students. It also improves behavioural outcomes such as teacher-student relationships and students' social acclimatisation and satisfaction at school, and reduces the level of violence and classroom disruption.

In England, a reform in the 1980s allowed state high schools to opt out of local authority control. These Grant Maintained Schools became autonomous schools funded directly by the central government – perhaps a precursor to today's Academies. Schools seeking autonomy had only to propose and win a majority vote among current parents. Almost one in three high schools voted on autonomy between 1988 and 1997, and using regression discontinuity design, Clark (2009) finds large achievement gains at schools in which the vote barely won, compared with schools in which it barely lost.

A programme in Pakistan offers the opportunity to study the quality assurance system for a public-private partnership programme. The Punjab Education Foundation offers public subsidies conditional on minimum learning levels to low-cost private schools. Schools are ejected from the programme if they fail to achieve a minimum pass rate in the test in two consecutive attempts, making the test high stakes. Sharp ***regression discontinuity*** estimates show that the threat of programme exit on schools that barely failed the test for the first time induces large learning gains. The large change in learning between the first two test rounds is likely to be attributable to this accountability pressure given that a large share of new programme entrants failed in the first test round (Barrera-Osorio and Raju 2010). Another study shows evidence of large positive impacts on the number of students, teachers, classrooms and blackboards in the classrooms of treated schools (Barrera-Osorio and Raju 2011).

Matching

When randomisation is not possible for whatever reason, but there is good information on student characteristics and backgrounds, then researchers can use the quasi-experimental method of creating similar groups by matching the observable information. This technique is used by Krueger and Zhu (2004) to re-examine data from the New York City school choice programme. In principle, random assignment would be expected to lead treatment status to be uncorrelated with all baseline characteristics. Students with missing baseline test scores, which encompasses all those who were initially in kindergarten and 11 per cent of those initially in grades 1-4, were excluded from previous analyses of achievement, even though these students were tested in the follow-up years. ***Propensity score matching*** was used to include students with missing baseline test scores, thus increasing the sample size by 44 per cent. For African American students, the only group to show a significant, positive effect of vouchers on achievement in past studies, the difference in average follow-up test scores between the treatment group (those offered a voucher) and control group (those not offered a voucher) becomes statistically insignificant. In addition, the effect of vouchers is found to be sensitive to the particular way race/ethnicity was defined. Previously, race was assigned according to the racial/ethnic category of the child's mother, and parents who marked 'other' and wrote in 'Black/Hispanic' were typically coded as non-Black and non-Hispanic. If children with a Black father are added to the sample of children with a Black mother, the effect of vouchers is small and statistically insignificant at conventional levels.

In 1999 the city of Bogota, Colombia launched the concession school programme designed to broaden the coverage and quality of basic education. It consists of a contract between a group of private school providers – selected from among the best schools in Colombia – and the public education system. The city built modern school buildings in 25 disadvantaged neighbourhoods such that private agents provide education for low-income students. Barrera-Osorio (2007) tests three main hypotheses concerning the impact of concessions on the quality of education: first, drop-out rates are lower in concession schools than in similar public schools; second, other public schools near to the concession schools have lower drop-out rates in comparison with other public schools outside the area of influence; and third, test scores from concession schools are higher than scores in similar public schools. The results of the ***matching*** study show significant effects on all three counts. Efficiency improves for concession school students and cognitive achievement rises; moreover, public schools in the vicinity facing competition from the concession schools also improve significantly.

Most schools in Chile provide both primary and secondary education. Lara, Mizala and Repetto (2011) analyse the effect of private voucher education on student academic performance. The researchers analyse the effect of private voucher education on students who are forced to enrol at a different school to attend secondary education once they graduate from primary schooling – structural switching. Moreover, the dataset the authors use contains information on previous academic achievement and thus allows them to identify differences in student characteristics. The authors used propensity ***score matching*** and the ***difference-in-difference*** estimation method and found that that private voucher education led to small, sometimes not statistically significant, differences in academic performance. The estimated effect of private voucher education amounts to about 4 to 6 per cent of one standard deviation in test scores.

Difference-in-difference/Phase-in

Retrospective evaluations with over-time data typically use difference-in-difference estimations. These quasi-experimental techniques sometimes also permit researchers to use programme phase-in to create counterfactual groups. In the case of England's academy schools, Machin and Vernoit (2011) use a school-level *difference-in-difference* method over time and compare specific schools that were state-maintained and eventually became academies as a control group: ***phase-in***. The introduction of academy schools into the English secondary school sector has allowed schools to gain more autonomy and flexible governance by changing their school structure. The researchers consider the impact of an academy school conversion on their pupil intake and pupil performance and possible external effects working through changes in the pupil intake and pupil performance of neighbouring schools. They focus on the period 2001 to 2009 and compare outcomes of interest in academy schools to a specific group of comparison schools, those state-maintained schools that go on to become academies after the sample period ends. This approach allows them to produce a well-balanced treatment and control group. Results suggest that moving to a more autonomous school structure through academy conversion generates a significant improvement in the quality of their pupil intake and a significant improvement in pupil performance. They find significant external effects on the pupil intake and the pupil performance of neighbouring schools. All of these results are strongest for the schools that have been academies for longer and for those who have experienced the largest increase in their school autonomy. Such benefits, however, take several years to materialise, something that is found in other autonomy reforms (Bruns, Filmer and Patrinos 2011).

Instrumental variables and Heckman selection models

Often, the lack of randomisation requires the search for other sources of variation to identify the programme. The standard instrumental variable (IV) estimator can identify the effect of the intervention only for the sub-populations affected by the observed changes in the instruments; sub-populations which respond most to changes in the instruments will have the largest effects on the magnitude of the IV estimate. Many stringent assumptions must be made before an instrument is seen as valid. IV estimates are inconsistent if the instruments are correlated with the error term in the equation of interest. Another problem is caused by the selection of 'weak' instruments, instruments that are poor predictors of the endogenous question predictor in the first-stage equation. Therefore, most IV studies come with many caveats and assumptions. This may explain the wide discrepancy of results obtained even for the same programme.

> An example: if researchers use distance as an instrument for attending a Catholic school in an education production function, then they are identifying the effect of Catholic schooling in the sub-population which would obtain a Catholic school education if a Catholic school is present but which would not obtain a Catholic school education if a Catholic school is not present. This empirical approach does not, without further assumptions, tell the researchers anything about the effect of Catholic schooling among people who would either always or never receive a Catholic school education regardless of whether a local Catholic school exists.

Chile was one of the first countries to implement a school choice programme for the stated purpose of improving efficiency in education. But as important as the 1981 programme was, the nationwide implementation presents two important difficulties for the purpose of attribution. First, participation was not randomised; therefore, it is difficult to disentangle the effects of the programme from the inherent difference in populations due to self-selection. Second, it is a universal programme, implying that the counterfactual of no vouchers is difficult to construct (Hoxby 2003b). Research on Chile has been subject to a high level of scrutiny. The literature presents mixed results, to say the least. Several early articles present data in which subsidised, private schools obtain higher standardised test scores than do state schools, though these studies are prone to selection bias (see, for example, Rodriguez 1988; Aedo and Larranaga 1994; Aedo 1997; Contreras 2001). A second batch of studies uses individual level information with large samples (Romaguera 1999; Mizala and Romaguera 2000; Gallegos 2002), though they too are prone to biases. Several studies advance the previous estimations, with strategies to overcome the problem of self-selection using the Heckman correction method (see, for example, Sapelli and Vial 2002, 2004) and IV approach (see, for example, Auguste and Valenzuela 2003; Hsieh and Urquiola 2006; Gallegos 2006). Long-term benefits of vouchers have been examined with IV estimates and show significant effects in the labour market (Bravo, Mukhopadhyay and Todd 2010; Patrinos and Sakellariou 2011).

Elsewhere instruments are used to assess various policies. These include 'Free Schools' in Sweden (Ahlin 2003; Björklund et al. 2004; Böhlmark and Lindahl 2008; Sahlgren 2010; Sandström and Bergström 2005), showing generally positive effects. They have also been used to assess 'Free Schools' in Denmark (Andersen 2005), showing generally no effects. Instruments have also been used to evaluate school choice in the Netherlands (Himmler 2007; Patrinos 2011), showing that vouchers increase academic achievement. Researchers have used the public funding of Catholic schools to document competition effects in the Canadian provinces of Ontario and Quebec (Card et al. 2010; Lefebvre et al. 2009), both showing significant and positive effects. Subsidising private high schools is shown to be effective in raising academic achievement in both Côte d'Ivoire (Sakellariou and Patrinos 2009) and Japan (Akabayashi and Arakia 2011).

6 Conclusion

Rigorous studies indicate that the non-state sector can improve learning outcomes of students. However the body of evidence is limited and more research is needed. The most convincing evidence relates to programmes that are well targeted towards disadvantaged students. But our knowledge base is informed by relatively few randomised control trials, especially outside of the United States.

A greater number of randomised trials in future research will increase our understanding of the types of interventions that improve educational outcomes. Randomised studies require fewer assumptions and reduce biases, thus allowing researchers to produce robust findings. While non-randomised studies offer important insights, they require more data. Ideally non-randomised studies require longitudinal data at the student level that can be traced back in time before the programme, tracking a significant number of variables.

In cases when a full or national pilot randomised trial cannot be undertaken, then the information from small-scale evaluations can be used to provide a case for roll-out and inform the design of future large-scale evaluations. Small-scale evaluation give a sense of the scale and time needed to see impact. This is crucial for setting up a more elaborate evaluation because it would clarify how long it might take for impact to occur and what sort of effects to expect. For instance, such a study could show by how much test scores will change over time.

In addition to evaluating the benefits of the programme itself – its internal validity – researchers could also consider spill-over/competition effects. How does the programme affect nearby schools? Are there spill-over effects? Does competition improve overall results? Moreover, beneficiary schools can be used to test specific hypotheses, such as peer and competition effects. Programmes may have differential effects on different sub-populations, for example: girls and boys, disadvantaged groups, rural residents, ethnic minorities. Thus, it is important to investigate differential, heterogeneous effects (Lamarche 2011).

It is important to ensure an impact evaluation is in place early in the intervention to capture the effect of the innovative approach. Retrospective evaluations can be used but are far more difficult to implement, are likely to have more biases, and more investment is needed in order to collate a higher amount of data to support the evaluation. Constantly changing political will is also another reason for ensuring the impact is captured and that policy decisions are evidence based.

Creating a culture of evaluation will ensure that all stakeholders understand that the evaluation results will be used to demonstrate impact. The use of any rigorous technique helps build acceptance of the concept and provides evidence to help support the innovation. It is, therefore, important to ensure all stakeholders understand the importance of impact evaluation going forward and their role in helping to support the evaluations. Stakeholders will include educators, administrators, school providers, managers, policymakers, parents, and the general public. Transparency and consultation should be used to strengthen buy-in to the impact evaluation.

Rigorously measuring impact enables evidence-based policymaking. The private sector could offer innovative solutions to improving educational outcomes, but it is the Government's responsibility to ensure their impact is effectively measured, so that those programmes that are effective can be expanded.

Impact evaluation can support best practice sharing between as well as within systems. Governments around the world are striving to improve educational outcomes. Impact evaluations can facilitate evidence-based best practice sharing. They can ensure policymakers are not introspective but look outwards to systems around the world for innovative proven solutions to improve educational outcomes for future generations.

References

Abdulkadiroglu, A., Angrist, J.D., Dynarski, S.M., Kane, T.J. and Pathak, P.A. (2011) 'Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots', *Quarterly Journal of Economics* (forthcoming).

Ahlin, A. (2003) 'Does School Competition Matter? Effects of a Large-scale School Choice Reform on Student Performance.' *Dep*artment of Economics, Uppsala University Working Paper No. 2.

Akabayashi, H. and Arakia, H. (2011) 'Do Education Vouchers Prevent Dropout at Private High Schools? Evidence from Japanese Policy Changes', *Journal of the Japanese and International Economies* (forthcoming).

Andersen, S.C. (2005) *Selection and competition effects in a large-scale school voucher system.* For presentation at the conference EPCS 1005, 31 March – 3 April at the University of Durham, UK.

Angrist, J.D. and Imbens, G.W. (1995) 'Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity', *Journal of the American Statistical Association* 90(430): 431–42.

Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A. and Walters, C.R. (2010) 'Inputs and Impacts in Charter Schools: KIPP Lynn.' *American Economic Review: Papers & Proceedings* 100(2): 1–5.

Angrist, J.D., Bettinger, E., Bloom, E., King, E. and Kremer, M. (2002) 'Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment', *American Economic Review* 92(5): 1535–58.

Angrist, J.D., Bettinger, E. and Kremer, M. (2006) 'Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia', *American Economic Review* 96(3): 847-862.

Athey, S. and Imbens, G.W. (2006) 'Identification and Inference in Nonlinear Difference-in-Difference Models', *Econometrica* 74(2): 431-497.

Auguste, S. and Valenzuela, J.P. (2003) *Do Students Benefit from School Competition? Evidence from Chile.* Ann Arbor: University of Michigan.

Barrera-Osorio, F. (2007) 'The impact of private provision of public education: empirical evidence from Bogota's concession schools.' World Bank Policy Research Working Paper No. 4121, Washington DC.

Barrera-Osorio, F. and Patrinos, H.A. (2009) 'An International Perspective on School Vouchers,' in M. Berends, M.G. Springer, D. Ballou and H.J. Walberg (eds), *Handbook of Research on School Choice.* New York: Routledge.

Barrera-Osorio, F. and Raju, D. (2010) 'Short-run Learning Dynamics under a Test-based Accountability System – evidence from Pakistan', South Asia Region Education Unit & Human Development Network Education Unit, World Bank Policy Research Working Paper 5465.

Barrera-Osorio, F. and Raju, D. (2011) 'Evaluating public per-student subsidies to low-cost private schools: regression – discontinuity evidence from Pakistan', World Bank Policy Research Working Paper 5638.

Bettinger, E., Kremer, M. and Saavedra, J. (2010) 'Are Educational Vouchers Only Redistributive?' *Economic Journal* 120(546): F204-F228.

Björklund, A., Edin, P. Fredriksson, P. and Krueger, A. (2004) 'Education, Equality and Efficiency – an Analysis of Swedish School Reforms during the 1990s.' IFAU Report No. 1.

Böhlmark, A. and Lindahl, M. (2008) 'Does School Privatization Improve Educational Achievement? Evidence from Sweden's Voucher Reform.' IZA Discussion Paper No. 3691.

Bravo, B., Mukhopadhyay, S. and Todd, P.E. (2010) 'Effects of school reform on education and labor market performance: Evidence from Chile's universal voucher system.' *Quantitative Economics* 1(1): 47-95.

Bruns, B., Filmer, D. and Patrinos, H.A. (2011) *Making Schools Work: New Evidence on Accountability Reforms.* Washington DC: World Bank.

Card, D., Dooley, M.D. and Payne, A.A. (2010) 'School Competition and Efficiency with Publicly Funded Catholic Schools.' *American Economic Journal: Applied Economics* 2(4): 150-76.

Clark, D. (2009) 'The Performance and Competitive Effects of School Autonomy.' *Journal of Political Economy* 117(4): 745-783.

Center for Research on Education Outcomes (CREDO) (2009) 'Multiple Choice: Charter School Performance in 16 States.' Center for Research on Education Outcomes, Stanford, CA.

Contreras, D. (2001) 'Evaluating a Voucher System in Chile: Individual, Family and School Characteristics.' Universidad de Chile, Facultad de Ciencias Económicas y Administrativas, Documento de trabajo no. 175.

Dehejia, R.H. and Wahba, S. (2002) 'Propensity Score-Matching Methods for Nonexperimental Causal Studies', *Review of Economics and Statistics* 84(1): 151-161.

Diamond, A. and Hainmueller, J. (2007) *The Encouragement Design for Program Evaluation*, IFC and Harvard University.

Dobbie, W. and Fryer, R. (2011) 'Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone', *American Economic Journal: Applied Economics* 3(3): 158-187.

Duflo, E., Glennerster, R. and Kremer, M. (2007) 'Using Randomization in Development Economics Research: A Toolkit,' in T.P. Schultz and J. Strauss, (eds), *Handbook of Development Economics*, Elsevier Science Ltd: North Holland, Vol. 4: 3895-62.

Epple, D. and Romano, R.E. (1998) 'Competition between Private and Public Schools, Vouchers, and Peer-Group Effects', *American Economic Review* 88(1): 33-62.

Friedman, M. (1955) 'Role of Government in Education,' in R. Solo (ed.), *Economics and the Public Interest.* New Brunswick, NJ: Rutgers University Press.

Gallegos, F. (2002) 'Competencia y Resultados Educativos. Teoría y Evidencia para Chile.' *Cuadernos de Economía* 39(118): 309-352.

Gallegos, F. (2004) *School Choice, Incentives, and Academic Outcomes: Evidence from Chile.* Boston: Massachusetts Institute of Technology.

Gallegos, F. (2006) *Voucher-School Competition, Incentives and Outcomes: Evidence from Chile.* Boston: Massachusetts Institute of Technology.

Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B. and Vermeersch, C.M.J. (2010) *Impact Evaluation in Practice.* Washington DC: World Bank.

Gertler, P., Patrinos, H.A. and Rubio-Codina, M. (2007) *Impact Evaluation for School-Based Management Reform* (Designing Impact Evaluation No. 10.) Washington DC: World Bank.

Greene, J.P., Peterson, P.E. and Du, J. (1999) 'Effectiveness of School Choice: The Milwaukee Experiment', *Education and Urban Society* 31(2): 190-213.

Hahn, J., Todd P. and Van der Klaauw, W. (2001) 'Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design.' *Econometrica* 69(1): 201-209.

Heckman, J. (1976) 'The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models', *Annals of Economic and Social Measurement* 5(4): 475-492.

Heckman, J. (1979) 'Sample Selection Bias as a Specification Error', *Econometrica* 47(1): 153-161.

Heckman, J., Ichimura, H. and Todd, P. (1998) 'Matching as an econometric evaluation estimator', *Review of Economic Studies* 65: 261-294.

Himmler, O. (2007) *The Effects of School Choice on Academic Achievement in the Netherlands.* Göttingen: Georg-August-Universität.

Hoxby, C.M. (2009) *A Statistical Mistake in the CREDO Study of Charter Schools* Stanford University, Stanford, CA.

Hoxby, C.M. (ed). (2003a) *The Economics of School Choice.* Chicago: University of Chicago Press.

Hoxby, C.M. (2003b) 'School Choice and School Competition: Evidence from the United States', *Swedish Economic Policy Review* 10: 9-65.

Hoxby, C.M. (2000) 'Does Competition among Public Schools Benefit Students and Taxpayers?' *American Economic Review* 90(5): 1209-1237.

Hoxby, C.M. and Murarka, S. (2009) 'Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement', NBER Working Paper No. 14852, Cambridge, MA.

Hoxby, C.M. and Rockoff, J.E. (2005) 'Findings from the City of Big Shoulders', *Education Next* 5(4): 52-58.

Hsieh, C-T. and Urquiola, M. (2006) 'The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program', *Journal of Public Economics* 90(8-9): 1477-1503.

Hughes, R.E. (1975) 'James Lind and the cure of scurvy: an experimental approach', *Medical History* 19: 342-351.

Kang, C. (2007) 'Classroom peer effects and academic achievement: Quasi-randomized evidence from South Korea', *Journal of Urban Economics* 61(3): 458-495.

Kim, J., Alderman, H. and Orazem, P.F. (1999) 'Can Private School Subsidies Increase Enrollment for the Poor? The Quetta Urban Fellowship Program', *World Bank Economic Review* 13: 443-65.

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

Ladd, H.F. (2002) 'School Vouchers: A Critical View', *Journal of Economic Perspectives* 16(4): 3-24.

Lamarche, C. (2011) 'Measuring the incentives to learn in Colombia using new quantile regression approaches', *Journal of Development Economics* 96(2): 278-288.

Lara, B., Mizala, A. and Repetto, A. (2011) 'The Effectiveness of Private Voucher Education: Evidence from Structural School Switches', *Educational Evaluation and Policy Analysis* 33(2): 119-137.

Lavy, V. (2010) 'Effects of Free Choice among Public Schools', *Review of Economic Studies* 77(3): 1164-1191.

Lefebvre, P. and Merrigan, P. (2009) *Private Schools Do Make a Difference on Math Scores: Longitudinal Evidence from a Canadian Experiment.* Montreal: Université du Québec à Montréal.

Lewis, L.A. and Patrinos, H.A. (2011) *Framework for Engaging the Private Sector in Education.* Washington, DC: World Bank.

Loeb, S., Valant, J. and Kasman, M. (2011) 'Increasing Choice in the Market for Schools: Recent Reforms and Their Effects on Student Achievement', *National Tax Journal* 64(1): 141-164.

Machin, S. and Vernoit, J. (2011) 'Changing School Autonomy: Academy Schools and their Introduction to England's Education', Centre for the Economics of Education Discussion Paper 123, London School of Economics.

Neal, D. (2002) 'How Vouchers Could Change the Market for Education', *Journal of Economic Perspectives* 16(4): 25-44.

Nechyba, T.J. (1999) 'School Finance Induced Migration and Stratification Patterns: The Impact of Private School Vouchers', *Journal of Public Economic Theory* 1(1): 5-50.

Nechyba, T.J. (2000) 'Mobility, Targeting, and Private-School Vouchers', *American Economic Review* 90(1): 130-146.

OECD (2011) *The High Cost of Low Educational Performance: The Long Run Economic Impact of Improving PISA outcomes.* OECD.

Patrinos, H.A., Barrera-Osorio, F. and Guaqueta, J. (2009) *The Role and Impact of Public-Private Partnerships in Education.* Washington DC: World Bank.

Patrinos, H.A. (2011) 'Private education provision and public finance: the Netherlands', *Education Economics* (forthcoming).

Patrinos, H.A. and Sakellariou, C. (2011) 'Quality of Schooling, Returns to Schooling and the 1981 Vouchers Reform in Chile', *World Development* 39(12): 2245-2256.

Rosenbaum, P.R. and Rubin, D.B. (1983) 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika* 70(1): 41-55.

Sahlgren, G.H. (2010) 'Schooling for Money: Swedish Education Reform and the Role of the Profit Motive.' IEA Discussion Paper No. 33. London: Institute of Economic Affairs.

Sakellariou, C. and Patrinos, H.A. (2009) 'The Equity Impact of Public Finance of Private Education Provision in Côte d'Ivoire', *International Journal of Educational Development* 29: 350-356.

Sandström, F.M. and Bergström, F. (2004) 'School Vouchers in Practice: Competition Will Not Hurt You', *Journal of Public Economics* 89(2-3): 351-380.

Sapelli, C. and Vial, B. (2004) 'Peer Effects and Relative Performance of Voucher Schools in Chile', Paper presented at the American Economic Association meetings, 3 January 2004, San Diego, CA.

Sapelli, C. and Vial, B. (2002) 'The Performance of Private and Public Schools in the Chilean Voucher System.' *Cuadernos de Economía* 39(118): 423-454.

Schweinhart, L.J., Barnes, H.V. and Weikart, D.P. (1993) *Significant benefits: The HighScope Perry Preschool Study through age 27.* Ypsilanti, MI: HighScope Press.

Shadish, W., Cook, T.D. and Campbell, D.T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA and New York: Houghton Mifflin Company.

Annex 1: Private sector engagement

- **Private schools:** schools that are not owned by Government and are financed privately, typically through fees (i.e. no subsidies)
- **Private funded schools:** privately owned and managed schools (by private sector, church, religious groups, NGOs, foundation, charities, etc.) which receive funding from the Government but funding is not outlined in a contract on a per-student basis
- **Private contracted schools:** private schools contracted by the Government, where the transfer of public funds depends on the school satisfying specific conditions
- **Private management schools:** a private organisation that operates and manages schools owned by the Government (e.g. Charters, Academies, Concession schools, etc.)
- **Market contracted schools:** public schools, private contracted schools or private management schools are implicitly contracted by the student such that the funding follows the student to the school of their **choice** (e.g. vouchers)

Summary

Type of school provision	Publicly funded	Contracted	Ownership
Private (No subsidies)	No	No	Private
Private funded	Yes	No	Private
Private contracted	Yes	Yes, Government	Private
Private management (Charters)	Yes	Yes, Government	Government
Market contracted (Vouchers)	Yes	Yes, Students	Mixed

CfBT Education Trust
60 Queens Road
Reading
Berkshire
RG1 4BS

+44 (0) 118 902 1000

www.cfbt.com

PMS 5958 01/12 ISBN 978-1-907496-73-8