What is the evidence of the impact of vouchers (or other similar subsidies for private education) on access to education for poor people?

Protocol written by Trevor Fronius, Anthony Petrosino, Claire Morgan, Learning Innovations at WestEd

EPPI-Centre
Social Science Research Unit
Institute of Education
University of London

April 2012

PROTOCOL

This protocol should be cited as: Fronius T, Petrosino A, Morgan C (2012) What is the evidence of the impact of vouchers (or other similar subsidies for private education) on access to education for poor people? Protocol. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

Contents

1. Background

1.1 Aims and rationale for review

It has been well established that education is a key to economic growth and social welfare in developing as well as developed nations. Research indicates that investments in education can yield returns in poverty reduction, improved health outcomes, and economic growth (UNESCO, 2007; Hannum & Buchmann, 2004; Herz & Sperling, 2003). In addition, increased access to education may contribute to increased political participation and more equitable sharing of economic and political power (Birdsall, 1999). Education for girls is particularly critical, as studies show that improvements in the infant mortality rate, child nutrition, and school enrolment are closely associated with maternal education (Birdsall, Levine, & Ibrahim, 2005; Herz & Sperling, 2003; World Bank, 2008). Yet, despite the importance of education to the developing world, more than 100 million primary school aged children are not in school, and of those that are, many—49 percent in Africa, for example—do not complete primary school (Birdsall, Levine, & Ibrahim, 2005).

Low educational attainment in the developing world is often attributed in part to the private costs associated with sending children to public and private schools (Filmer & Pritchett, 1998; Filmer & Pritchett. 2001; Bentaouet Kattan and Burnett, 2004; Yardley, 2005). Direct and indirect educational costs may also exclude low-income households from private schools or better quality public schools. Direct costs may come in the form of school fees and textbooks, and indirect costs represent, for example, costs due to lost labour opportunities for children. Children of the poorest income quintile in developing countries consistently represent the lowest percentage of eligible students attending school (Thapa & Mahendra, 2010).

The limited and unbalanced availability of affordable schooling in the developing world can lead to overcrowding and poor educational quality for children that arguably need the most help (Angrist, 2002). Lack of resources and access to quality education may very likely perpetuate the educational inequalities within developing countries. It is one goal of voucher programs in developing nations, and other similar incentive-based programs, to address these inequities.

A general education voucher system" in the broadest sense is a payment made by the government to a school [or directly to the parent] chosen by the parent of the child being educated; the voucher finances all or most of the tuition being charged. The system introduces competition among public schools and between public and private schools; it enables schools to offer diverse educational packages to meet the different preferences of parents" (West, 1997, pg. 83). *It is important to note that this review focuses exclusively on general voucher programs, as defined above. This review focuses on government-sponsored voucher programs that incentivize schools to improve education.* There are incentive-based models that do not necessarily offer school choice, which is typically the emphasis of voucher programs. For example, many developing (and industrialized) nations have adopted various incentive-based programs to promote primary and secondary education. Colombia, Bangladesh, Brazil, Mexico, Ecuador, Pakistan, and Chile are examples of nations that have implemented various types of incentive-based programs (e.g., conditional cash transfer [CCT], food for education, stipends) typically targeted at marginalized populations to promote enrolment and attendance of school-age children (Angrist, 2002; Ravallion & Wodon, 1999;

Behrman et al., 2000; Hsieh & Urquiola, 2003; Barrera-Osorio, 2007; Kim et al., 1998; Schady,2006).

In general terms, a voucher model provides government funds to schools based on enrolments or toeligible families to apply toward the tuition at participating private or public schools. Supporters claim that a major outcome is to increase school choice by alleviating financial burden associated with private (or public) school enrolment costs, which ultimately facilitates school competition, not only amongst private schools but also among public schools (who now have to compete for their students), thereby increasing education quality for all students (e.g., Friedman, 1955). Although voucher programs vary widely across nations, the goals are somewhat similar: to allocate sufficient funds to increase school enrolment, to increase access to quality education, and to eventually improve public school education (McEwen & Carnoy, 2000).

Several countries have adopted voucher programs and have reported increases in private school enrolment (e.g., Angrist et al., 2002). For example, in Chile, private voucher school enrolment increased from 15% to approximately 33% of total enrolment, between 1981 – the voucher scheme inception – and 1996 (McEwen & Carnoy, 2000). Bangladesh also saw female enrolment increase by nearly 65 percent during its program pilot study (King et al., 1997). An evaluation of Columbia's program showed a 15 percent increase in private school enrolment among voucher awardees and statistically significant improvement among females for educational attainment (Angrist et al., 2002).

However, some argue that increases in enrolment are often accompanied by tradeoffs in educational quality as class sizes rise and overpower existing capacity in school infrastructure and staffing. Shafiq (2010) suggests that although there has been rapid growth in low-cost non-public schools in developing countries, there is little scientific evidence of the impact of educational vouchers specifically targeted at the poorest populations in the program locality. The evidence thus far validates the need for further examination of general (that is, all children are eligible) and targeted voucher programs and a concise understanding of program evaluations to date. Likewise, government funding for the schools may decrease, which may have counterproductive effects and lead some parents to be disillusioned about the education possibilities for their children and for some youth to drop out (World Bank, 2009a). In addition, financial instability and/or lack of true governmental commitment may threaten the sustainability of the voucher policy (Bentaouet-Kattan, 2006; World Bank, 2009a). Furthermore, an educational voucher program, or similar educational reform program alone, may not be enough to extend access to the poorest and most marginalized groups facing barriers such as high opportunity costs or transportation costs, poor health, and social conflict (Bentaouet-Kattan, 2006; World Bank, 2009b). It is critical, whenever possible, to analyze impact across socioeconomic status (SES) levels, regardless of overall World Bank classification (i.e., Low-income, Lower-middle income, Upper-middle income).

Moreover, vouchers may not pay enough for poor households to send their children to the better schools, and quality schools, whether public or private, are not available in all areas. Thus, the availability of vouchers may not translate into improved educational outcomes for children and in some cases may even increase inequality. For example, voucher schemes could incentivize schools to enrol the highest performing students, and to the extent that performance is correlated with social class, it would result in further inequities.

Although surges in private school enrolment following the introduction of education voucher programs in many developing countries have been reported, as described

above, the impact of voucher programs, exclusive to developing countries, has not been systematically reviewed (McEwan, 2000; Angrist et al., 2002; King et al., 1997; Thapa & Mahendra, 2010). This systematic review is unique and contributes to existing literature on vouchers (McEwan, 2000), by focusing on developing countries and drawing from new evaluation research conducted in the last twelve years. With the substantial monetary allocations provided to these programs, it is imperative to understand what works, what does not, and what steps for improvement may be made moving forward. In our information retrieval process for an ongoing systematic review of school enrolment policies in developing nations, we have identified a few studies relating to voucher programs in developing countries, with examples of randomized controlled trials (RCT) or rigorous quasi-experimental designs (QED). For the present review, more extensive searching, including hand searches, examining grey literature, and contacting experts in the field, will be used to ensure that we identify the rigorous impact studies that have been conducted on this topic. These studies will be used in analyses or discussions about estimates of program impact.

But such a review needs to go further. We also propose to use extensive search methods to identify a wider range of non-experimental and qualitative evaluative research studies. This literature will not be used to estimate program impacts, but to shed light on the important contextual issues relevant to vouchers in developing nations.

A systematic review of rigorous impact studies and non-experimental evaluations and qualitative research is essential for understanding the nature and quality of evaluation evidence in an area receiving attention by development agencies, non-governmental organizations (NGOs), and academic institutions. The analysis will address the wide variety of subsidies attached to voucher programs, along with the scale and allocation of funding. We believe a systematic review that focuses specifically on the evidence for application of education voucher programs in developing nations would be of import to a broad range of stakeholders, including those from governments of developing and donor nations, international and non-governmental bodies, and researchers.

1.2 Definitional and conceptual issues

1.2.1 Conceptual framework

The review draws its conceptual framework from the growing body of literature on demand-side education financing programs--particularly on incentive-based models (e.g., King et al., 1997). In its most general sense, in a voucher program, the government provides resources to households [or schools], to offset the costs at eligible primary and/or secondary schools. Vouchers, in essence, aim to manage current poverty, by lessening the financial burden of primary and/or secondary school attendance for households, and ease future poverty, by increasing human capital through higher educational attainment of children from poor households, leading to increased future earnings (West, 1997).

The educational voucher scheme was initially conceived by the economist, Milton Friedman (1955), as a way to foster competition amongst schools, which in turn should increase overall educational quality or eliminate those schools that cannot keep pace. This competition along with the voucher will increase school choice for all children (Levin, 2002). West (1997) explains the conceptual framework of voucher models as being comprised of four principles: (1) consumer choice, (2) personal advancement or opportunity to choose, (3) the promotion of competition, and (4) equal opportunity. West (1997) explains that the last principle, equal opportunity, is the logical outcome of the first three principles. It is often

represented in voucher models as the end result of increasing access to private schools and theoretically higher quality education. Equal opportunity is exemplified in vouchers that target poor households, by providing them with increased school choice. Some may argue that unless the program is specifically targeted to poor households, competition will not ensure equal opportunity. Levin (2002) expands on this idea with a comprehensive school voucher framework.

Levin's (2002) framework also focuses four criteria of voucher programs: 1) freedom to choose, 2) productive efficiency, 3) Equity, and 4) Social Cohesion. He notes that all voucher programs are not the same, nor are the overarching objectives. Voucher programs may place particular emphasis on one criterion over another by using three "policy instruments (Levin, 2002, p. 14): 1) finance, 2) regulation, and 3) support services; these are what defines and distinguishes individual voucher programs. Although some, as with Friedman's (1955) original proposal, focus mainly on freedom of choice and productive efficiency, advocates of targeted vouchers propose that placing emphasis on equity, social cohesion, and freedom of choice, will provide the greatest good for those who need it most (Levin, 2002).

We distinguish between conditional cash transfer programs and other payments that are made to households to promote enrolment and attendance from general voucher programs. A larger review that focuses on all payments, regardless of whether they promote school choice, is beyond the scope of this current review. This review, therefore, will isolate voucher models based on the criteria below.

1.2.2 Definitional Concepts

This review will identify and code studies that test whether providing voucher schemes will increase access to private or public education for poor households, as demonstrated by subsequent increased enrolment, as well as persistence, in school. We will also measure "spillover" effects, if reported in the original documents, on public and non-voucher schools in the same locality. In addition to school enrolment and student persistence measures, we will also code equity outcomes, student achievement and other school quality outcomes, such as cognitive and non-cognitive test scores, teacher and student attendance, and pupil-teacher ratio, when available. We rely on the following definitions to guide us in our definition of what constitutes an eligible voucher study.

Voucher: As aforementioned, an education voucher system" in the broadest sense is a payment made by the government to a school [or directly to the parent] chosen by the parent of the child being educated; the voucher finances all or most of the tuition being charged. The system introduces competition among public schools and between public and private schools; and it enables schools to offer diverse educational packages to meet the different preferences of parents" (West, 1997, pg. 83).

Similar subsidies: Other subsidies will be considered in the review. For example, we will consider "chits" and tax credits for primary or secondary education. Chits are provided to parents and represent a voucher or receipt that can only be cashed by designated schools. The schools collect chits and then redeem them for cash from an appropriate government authority, which they generally use to pay school staff salaries. Tax credits are most often applied to higher education, but studies testing such credits may be included in this review if their core function is similar to vouchers and they are implemented in primary or secondary schools. One issue with tax credits is that "poor" households typically cannot afford the upfront costs associated with a private education. Thus, even if there is a tax credit, it is unlikely that the household will have the means to support a private

education without an upfront stipend (West, 1997). We will also include vouchers for non-tuition expenses.

Private education: The term private education generally refers to private non-subsidized schools that charge tuition to all students. These may include NGO schools, madrasas, and religious schools. Although some public schools also charge a fee or tuition to students, generally speaking, these schools are also subsidized by public funds. This review will include, however, those studies that tested voucher programs that provide an open choice (households may choose public or private schools systems) or restrict choice to public schools.

Education: Education level, as it pertains to this review, refers to Kindergarten through grade 12. This may include primary schools, secondary schools, both, or the localized equivalent of either level. This review will not include higher education institutions. Although we specifically define private education, as presented in the review question, we will also include "spillover" effects in all K-12 schools (i.e., public, non-voucher, government, municipal, NGO) within the same locality.

Poor People: For the purpose of this review, we will focus on countries that are low-income, lower middle-, or upper middle-income based on the World Bank classification. The review is focused on access to education for "poor people." This review will include all voucher models, as defined above, and will note schemes that specifically attempt to target recipients based on socioeconomic status, limit the locality (if it is relative to poverty) of the program, use any type of means tested model, or scale voucher awards based on income if it is an unrestricted model. It is of particular interest if vouchers that target the poor have greater positive outcomes than flat rate voucher schemes. We will rely on the author(s)' definition, formal or informal, of socioeconomic status among program participants. Some primary studies do identify the program population. Even in Middle Income nations like Chile and Colombia, there is value is looking at how vouchers affect outcomes, and particularly if those outcomes are examined in the primary documents for the poorest socioeconomic subgroups.

1.3 Policy and practice background

1.3.1 Voucher programs

Voucher programs provide payments to a family or directly to a school on behalf of the family, to cover or offset the cost of school enrolment. In voucher programs, students frequently can choose private or public school enrolment usually from a cohort of schools selected for participation by the program. The goal of voucher programs is to create school competition, which should improve education quality across all program schools. In Chile, arguably the most well-known voucher system in the developing world, has been in place since the early 1980's. The Chilean voucher system decentralized the municipal school system and began financing both public and private school through its voucher program (McEwan et al., 2000). The program is publicly funded and it allows parents to select from public or private schools for their children.

In fact, three types of schools make up the Chilean educational system: municipal, private subsidized, and private non-subsidized, and the government monitors the educational quality at all three types to varying degrees (Behrman, et al., 2010). Municipal and subsidized private schools may charge an additional small tuition, but both are mainly financed through the voucher system. Private non-subsidized schools are financed solely through private tuition. Families can choose any school, but private schools may have a more selective admissions process, whereas

municipal schools can only be selective if demand exceeds supply (Behrman et al., 2010).

The Chilean model is an example of an unrestricted model. On the other hand, the Colombian model is considered a targeted voucher scheme. In Colombia, the Ministry of Education created a targeted voucher system (known as PACES) aimed at shifting students from crowded public secondary schools to more vacant private secondary schools. The program targeted students from poor families. Eligible households could enter a lottery[1] to receive a school voucher if the child was enrolled in a public school and was accepted to a private school prior to being awarded the voucher. Today, the voucher system--which originally covered tuition costs at the average priced private school--now only covers about half the cost of tuition (Angrist et al., 2002).

1.4 Research background

There is a relatively large amount of literature on voucher programs in general, but the evaluative research, outside of Chile and Colombia, appears less extensive. As previously mentioned, the current research on the topic, in general, offers two potential outcomes at odds with one another. At one end of the spectrum, it is theorized--in the most general sense--that providing vouchers, regardless of conditional terms, will alleviate the financial burden for poor households associated with sending their child to school (e.g., Sapelli, 2010). Moreover, because the voucher money follows the student, competition for students will force poor quality schools to either improve or close down. On the other hand, critics argue that the rising demand for education associated with voucher programs may lower education quality or isolate the poorest populations in the lowest quality school systems, thus exacerbating inequality. Poor households may not be able to "top-up" their voucher to afford the best quality schools because vouchers are often not equal to the full cost of tuition at non-subsidized private schools (e.g., Sapelli, 2010). In addition, voucher schools may select the most advantaged students, thus increasing inequality.

Advocates for voucher programs stress that in low-income countries, private enrolment is generally more desirable. In fact, "private enrolment as a proportion of total enrolment is 2-3 times higher than in industrialized nations", because of the typically low-quality education available in most public schools (Angrist et al., 2002, p.1535). High teacher and administrator absenteeism exacerbate low quality, which lead to many families removing their children from school (Gauri et al., 2003). Of course, not all private schools offer a higher quality of schooling than that found in public schools, which also vary greatly in quality.

The Chilean voucher system offers vouchers to all school-aged children. In 2009, the program passed a law to increase subsidies based on low SES, a positive, although some argue temporary, improvement from the program's original flat voucher model (Sapelli, 2010). Households can choose to send children to any school, limited only by relative proximity and ability to pay any additional school fees. The system produced a mass shift from public schools to voucher-eligible private schools and increased private school attendance almost 35 percent in the first 15 years of the program (Arenas, 2004). However, the overall achievement level for voucher private schools versus public schools has not been conclusively established (Contreras, 2002; Sapelli et al., 2005; Arenas, 2004).

---

[1] At the onset of PACES, vouchers were awarded to more than 125,000 students. In areas where the demand exceeded supply such as Bogota, a lottery system was used to determine voucher awardees.

Unlike the Chilean voucher scheme, Colombia never saw a substantial shift from public to private schools, mainly because most parents viewed the public schools positively. Moreover, as the voucher lost its relative value, the capital lost outweighed the benefits of private school attendance (Angrist et al., 2002). However, in a series of follow-up studies, voucher winners demonstrated significantly higher test scores, lower repetition rates, and higher secondary attainment (Angrist et al., 2002). A longer-term evaluation showed that voucher winners were also more likely to graduate from high school, take the college entrance exam, and score higher than lottery losers (Angrist et al., 2006).

While the results from Colombia appear promising, the differences between the Chilean and Colombian systems must be emphasized: open versus targeted; national versus selective localities; and differences in public school quality. While Chile and Colombia have implemented the two most widely known voucher programs in the developing world, there are a handful of countries that have also implemented voucher programs on a smaller scale (Braun-Munzinger, 2005).

To our knowledge, a systematic review of studies evaluating the impact of "general voucher" programs on poor people's access to private education has not been conducted. This review will identify rigorous experimental and quasi-experimental impact studies. If available, studies will also be identified that address the longer-term impacts of voucher programs, such as is the case with Colombia (Angrist et al., 2006).

While the goal is to quantitatively synthesize all eligible studies, a meta-analysis may not be possible or appropriate. There is still value added, however, even if a small yield of studies is identified. Such a review would document the need for more primary studies of voucher programs in the developing world (Thapa & Mahendra, 2010).

1.5 Objectives

The objective of this project is to systematically collect and analyze relevant, high quality evaluative research that responds to the question: What is the evidence of the impact of vouchers (or other similar subsidies for private education) on access to education for poor people? We will assess the breadth and quality of the literature base and then synthesize all available empirical evidence of the impact of providing vouchers including, but not exclusive, to the following types: (1) conditional; (2) unconditional/unrestricted; (3) targeted; (4) follow the child (i.e., governments subsidize "schools of choice" in strict proportion to enrollment); and (5) means-tested (in which children from families below a certain income level are eligible). We will examine the impact of these studies on access to education for poor people in developing nations (as defined by the World Bank). Policies can be implemented at the universal, community/district/targeted group, the grade/age group, or at school level. We will report on all relevant educational outcomes, including:

- Private/public school enrolment
- Equity (i.e., along Gender/SES lines) in access
- Student achievement and attainment outcomes (e.g., persistence, completion, repetition, dropout, and test scores)
- School/educational quality outcomes (e.g., teacher/student absenteeism, pupil-teacher ratio, infrastructure adequacy, and per-pupil expenditures)

- "Spillover" effects on public and non-voucher schools in the same area

To shed light on important mechanisms and contextual issues, we will also examine available non-experimental evaluation studies and qualitative evaluative research studies relevant to vouchers. The issues highlighted in these studies will be discussed in narrative fashion.

We will code and assess each source in accordance with the procedure discussed in section *2.2.1 Defining relevant studies: inclusion and exclusion criteria*.

2. Methods used in the review

The methods in place for our meta-analysis are consistent with those approved by the Campbell Collaboration and International Initiative for Impact Evaluation (3ie). The design, according to these standards should include: 1) clear inclusion/exclusion criteria; 2) an explicit and transparent search strategy; 3) systematic coding and analysis of included studies; and 4) (if possible) a meta-analysis (Campbell Collaboration, 2011). The goal of this systematic review is to gather sufficient data for a meta-analysis. In the event that there is a dearth of sufficient data, the reviewers will undertake a qualitative systematic approach to the review.

2.1 User involvement

2.1.1 Approach and rationale

In developing this protocol, we have consulted with education policy staff at DFID and with Najeeb Shafiq, a content area expert from the University of Pittsburgh. As a goal to reach any potential stakeholders and users of the review, we plan to present the findings at conferences of international educational organizations such as the American Educational Research Association (AERA), the American Evaluation Association (AEA), and the Comparative and International Education Society (CIES). We will pursue publication of the review in a peer-reviewed journal such as the *Journal of Development Effectiveness*, *World Development*, *Journal of Development Studies*, *Journal of Multidisciplinary Evaluation*, or *International Journal of Educational Development*. We will also share the review with any development organization (e.g., World Bank; UNESCO; UNISEF; SFAI; IDRA) that is funding programs or shaping policy in the school voucher area. Additionally, we will attempt to contact and share our review with any key authors identified during our searches. Finally, we will work closely with DFID to disseminate findings to any other key players in this area, through meetings, events, conferences, and publications.

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

For this review, we will only include studies that fit the following criteria:

(1) *The program evaluations have to be conducted in a country classified as a low income, lower middle, or upper middle income country* as defined by the World Bank at the time the intervention or program was implemented.

(2) *The evaluations must include outcomes that measure the impact of school vouchers (*as defined in section *1.5) on access to schooling.* The outcomes will be identified based on the level of assignment for the voucher (e.g., households, schools, students). We will collect all relevant outcome measures reported in the study, such as enrolment, student achievement and attainment outcomes, and school quality outcomes. Additionally, we will identify and summarize any cost-benefit information available in the report.

(3) *The evaluation study report is published or available between January 1991 and November 2011, without regard to language or publication type.* We will attempt to find English and non-English studies. In addition we will include both published and unpublished documents (e.g., dissertations, working papers, technical reports, conferences papers).

 (4) *Randomized controlled trials (RCTs) or quasi-experimental evaluations with some evidence that the groups being compared are equivalent.* Our review will include evaluations that randomly assigned entities (at any level) to intervention or control conditions. In our work for a previous review (Petrosino et al., 2009), we found that the control or comparison group typically receives no intervention or continues to receive the same intervention ("business as usual"). Due to the fact that randomization is simply not possible for all studies (e.g., retrospective evaluations, post-hoc evaluations), we will also

include quasi-experimental designs that provide evidence that the comparison groups were equated (e.g., regression discontinuity design [RDD] with analysis around the cut-off score, propensity score matching [PSM] designs, covariate matching designs, and other pre or post intervention matching). All studies that use randomization, regression discontinuity, propensity score matching, difference-in-difference (DID) methods, instrumental variable (IV) analysis, etc., are initially eligible (before and after designs, or non-equivalent comparison group designs are excluded from impact analyses). We will look specifically at baseline data on enrolment and socioeconomic status as critical indicators for group equivalence. The review team will use a rating system that scores group equivalence; only those studies that are rated as being reasonable to conclude that the groups were equivalent will be included. We will explore the impact of study design as a moderator in our later analyses (*see section 2.3.2*).

Although RCTs and QEDs with group equivalence will inform the effect size estimates, we will also include the following designs to further illuminate the conceptual discussion of school vouchers and shed light on voucher policy implementation issues:

(5) *Quasi-experimental without pre-test group equivalency studies* will be examined, but not included in effect size estimates.

(6) *Descriptive quantitative and qualitative studies.* To be included in the review, non-causal studies should identify themselves as an evaluative study (e.g. impact study, assessment, evaluation) and must provide sufficient methodological detail to be in some degree replicable—that is, they describe the data analyzed and clearly outline the specific methods employed to evaluate the vouchers program. These studies will not be included in estimates of effect sizes.

2.2.2 Identification of potential studies: Search strategy

Search strategy

Our goal is to identify both published and grey literature. Many of the databases in (1) below include grey literature (e.g., ERIC). The British Library indexes conference proceedings and makes these available in its "Integrated Catalogue." Web searches will often identify literature that is made available at websites but not published in journals. Our contact with colleagues is designed to get at more of the grey literature. To accomplish our goal, we will use five major strategies:

(1) Electronic searches of bibliographic databases. We will use available online resources and databases at WestEd and the University of Pennsylvania, including *British Education Index, PAIS International/Archive,* and *Sociological Abstracts.* See Appendix A for additional databases that will be searched. We will also use the "advanced search" options in Google for broad searches of the World Wide Web.

(2) Hand searches of relevant journals. Because electronic searches often miss relevant studies, we will hand search the table of contents, and the abstracts when necessary, of all issues of the journals most likely to publish studies on this topic. From our initial searches, the five journals that we have identified for hand search are: *Economic Development and Cultural Change, International Journal of Educational Development, Journal of Development Economics, World Bank Research Observer,* and the *World Bank Economic Review.* We may also explore *Comparative Education Review, Education Economics, Economics of Education Review, Educational Evaluation and Policy Analysis,* and *Journal of School Choice.*

- *(3) Specific examinations of online holdings of international development organizations and research firms.* This would also include international or national/federal agencies that either conduct or would be aware of possibly relevant evaluations in developing nations, including the World Health Organization, the U.K. Department for International Development (DFID), and the

United States Agency for International Development (USAID). It would also provide coverage of websites with great relevance to international development, including the Network for Policy Research, Review and Advice on Education and Training (www.norrag.org). Research firms such as RTI International (particularly its international education division at http://www.rti.org/page.cfm/International_Education) and the Academy for Educational Development (www.aed.org) will also be part of this search strategy. Other key databases with relevant holdings include: 3ie Data base of Impact Evaluations; Africabib (i.e., Africana Periodical Literature; African Women; African Studies Abstracts Online); Association for the Development of Education in Africa (ADEA) Publications and Documents; British Library for Development Studies catalogue; Consortium for Research on Educational Access, Transitions, and Equity (CREATE) bibliographic database; and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) Papers and Theses.

(4) Citation chasing. The reference section of every report that meets the inclusion criteria will also be visually scanned to determine whether any citations are possibly eligible evaluations. We will also check for relevant citations in background resources on the review topic. As noted in the eligibility criteria, we are not exclusively seeking English language reports. We will ask our colleagues from other nations for help in identifying any non-English studies. WestEd also has employees bilingual in Spanish, French, Japanese, Portuguese, and Chinese who can translate abstracts or full text documents in non-English to determine their eligibility for this review.

(5) Contacting the network of researchers in this area. There is a network of researchers that are conducting or are aware of experimental and quasi-experimental studies relevant to school vouchers in developing nations. We will identify the lead authors of such studies or relevant documents (e.g., reviews, non-evaluative studies), identify their emails from a Google search of the World Wide Web, and email them query letters. A full listing of all persons contacted will be included in the final review.

As previously mentioned, searches of these sources will be limited to any report published or available between January 1991 and November 2011.

We have established a bibliographical reference database to maintain a log of all included and excluded studies. The log includes a field that allows the reviewers to document the reason for exclusion.

2.2.3 Screening studies: applying inclusion and exclusion criteria

Search methods will identify a large number of citations and abstracts. Many of these will be easily excluded as not being relevant to the proposed review. In some cases, however, they will identify potentially eligible studies. Two persons will independently review all abstracts and identify studies that should be retrieved, i.e., is a potentially relevant study. If so, the full text documents of those potentially eligible studies will be retrieved and screened before the study can be formally included in the review. Fortunately, with the advent of the Internet and full-text electronic journal access, we will be able to rapidly retrieve the reports to do a more thorough reading. When a full text report is received, we will check to ensure that it includes randomization or quasi-experimental equating of study subjects and evaluates an intervention that provides some type of school voucher and includes data on private school enrollment. Two authors will review studies proceeding to this second stage independently, and disagreements will be resolved by consensus among all three authors. Any study that fits the scope of the review, but does not fit our definition of high-quality design, will be noted as such in our bibliographic database.

As mentioned in section *2.2.2,* we have developed a database that will allow us to track the screening process. Any potentially eligible studies will be included in the database, and we will document studies that are included or excluded at first or second stage screening and, if applicable, document the reason for exclusion, along with any inter-reviewer disagreements.

2.2.4 Characterising included studies

Studies included in the review will be identified by overall methodology (e.g., RCT, QED, descriptive) and by specific methodology (e.g. RDD, PSM, difference-in-difference, covariate matching). In addition, we will code studies by specific type of voucher model (e.g., conditional, unconditional, open, targeted, means-tested) and by level of assignment (e.g., national/universal voucher program, city voucher program, targeted by SES). We will include other studies not meeting the methodological screening in the report, but these studies will not be included in any effect size estimations.

2.2.5 Identifying and describing studies: quality assurance process

To ensure that we achieve acceptable coding reliability, we will have two researchers read and record information from all reports independently. We will assess coding reliability (i.e., inter-rater agreement) by using the percentage of agreement for each item, rather than reporting a global inter-rater reliability statistic. This will avoid inflating reliability measures with study characteristics that generally achieve perfect agreement (e.g. year of publication) with those that do not. Items with lower rates of agreement (less than 80%) will be investigated to determine the source for conflict. The authors will meet to resolve disagreements and discussing coded items, with the third author resolving any persisting differences. We will drop those items from our database in which resolution could not be reached, as well as items that lack clear interpretation.

2.3 Methods for synthesis

We have designed a coding instrument to guide us in the extraction of relevant information from each included impact study (Appendix 2.5). The coding instrument will capture some data in open-ended form that we may later collapse, if necessary, prior to our analyses. For example, we will employ an open-ended response to describe the treatment condition. Upon completion, we will collapse this variable to capture the most comment types and code all other responses as "other".

2.3.1 Assessing quality of studies

Impact evaluations

For impact studies that are the focus of the review and for which effect sizes will be reported, we will extract information about the assignment of units, and other methodological aspects of the evaluation. We will capture the level of assignment (e.g., national/universal, district, city, village) and whether the authors employ any subgroup analyses. In order to determine the overall study quality, we will focus on the following:

How the groups were equated and whether any problems with equating were reported. The integrity of a randomized experiment or a quasi-experiment largely rests on how faithfully the equating procedures were implemented. We will code information about randomization and the quasi-experimental matching or equating procedures that were used in the study. In randomized experiments, this includes how much of the originally randomly assigned sample actually received the treatment (slippage from the "intention to treat" sample). We will code this information using a two-stage process. The first stage is a more detailed gathering of the facts about the assignment. The second stage will be comprised of a rating that will indicate the degree to which group equating was compromised by any reported problems. We first make any statistical considerations (i.e.,

equivalence indicators; low statistical power; large sample population), and then assign a rating for issues related to group equivalence as "Low", "Medium", or "High".

We will remove the pretest/baseline effect size from the overall effect size (difference-in-difference) when appropriate and possible to do so. We will create an effect size, when possible, for the difference between baseline groups and subtract this from the effect size between groups at follow-up. We are employing these corrections in our overall study of school enrolment interventions and will use them here.

Whether the researchers report a loss of participants from the initial assigned sample at the end of the study, how much attrition is reported, and whether the attrition differentially affects one group or the other. Such attrition, if it is significant, can compromise the equating of groups, particularly if different types of people drop out from the intervention than dropped out from the other conditions. We will code specific information on the amount of attrition (if it occurred) and whether it was differential in nature. We also propose a two-stage coding of attrition. First, we will code, in descriptive and qualitative form, any implementation problems noted by the investigators. Second, we will then rate the degree of attrition problems (with the standard being how the attrition problem affects a "fair test" of the program under investigation) as "high," "moderate," or "low." As a further methodological check, we will conduct sensitivity analyses to examine if results change based on dropping studies that experienced different levels of attrition (10%-25%; 25%-50%; 50% or more), or if results change if studies that report differential attrition (significant losses from one group or the other, with a difference in attrition of 5% or more between the groups). These are obviously subjective classifications, but the goal is to determine if the attrition compromised the study findings. We will collect any contextual information that informs reasons for attrition and include in a comments section or study summary.

Whether the program experienced significant implementation and fidelity problems. The first two issues deal with the implementation of the evaluation. This issue deals with the implementation of the program; there may be no observable program impact because no "real program" was ever implemented. We propose two-stage coding of implementation. First, we will code, in descriptive and qualitative form, any implementation problems noted by the investigators. Second, we will then rate the degree of implementation problems (with the standard being how the implementation problem affects a "fair test" of the program under investigation) as "high," "moderate," or "low."

Other evaluative studies

For the other evaluative studies located in the searches (quasi-experimental designs without equating of groups, descriptive quantitative, qualitative) two authors will also rate the degree of threat to the validity of the conclusions as "high", "moderate", or "low" based on our appraisal of such issues as the appropriateness of the methodology and data used, how explicitly the methods are detailed, whether the program experienced significant implementation or fidelity problems, the rigor of the data analysis, and the adequacy of the discussion of the findings.

2.3.2 Overall approach to and process of synthesis

2.3.2.1 Selection of studies for synthesis

As previously discussed, this review will synthesize randomized and high-quality quasi-experimental evaluations that detail efforts that meet acceptable rating for equating treatment and control groups. The process for synthesizing this data is discussed later [*see section 2.3.2.4*]. Although the following data will not be used directly in the effect size estimates, each will be coded and reported in the review:

Treatment of descriptive research
Descriptive quantitative and qualitative research will be synthesized to map the extent, types, and quality of the evidence base in the topic area and to provide contextual background to inform interpretation of the causal studies. In addition, qualitative data from the experiments and quasi-experiments will be used to illuminate three particular areas: (1) the context for the intervention; (2) the theory or mechanisms by which the program is supposed to impact the ultimate outcomes; and (3) the quality and nature of the intervention and comparison condition.

Treatment of economic data
We will report on any economic data included in the primary studies that are included in the review. This includes information on the costs of the program, any analysis of the cost-effectiveness of the intervention (e.g., the cost per child enrolled) and cost-benefit studies (e.g., the sum costs and benefits of the intervention). It is important that this information be linked in some way to the primary outcome studies so that it can be retrieved and coded appropriately.

2.3.2.2 Key coding variables

For the descriptive quantitative and qualitative studies for which effect sizes will not be reported, we will complete a structured abstract to capture information from the studies in the following categories: research questions, methods, location, participants, and findings.

The impact evaluation coding instrument (Appendix 2.6) discussed above will be our primary vehicle for recording and quantifying information from each impact study. We will capture any data that can be categorized or quantified, and which may be relevant to the current review. Any additional data (e.g., contextual, qualitative, summary information) will be recorded in a comments section, which will be used to draft short study summaries.

The coding instrument will capture data in the following areas:

Researcher and Study Characteristics:
Study reports can be used to provide information about the publication and characteristics about the experiment and the context. For example, we will extract data about the type of publication the study was reported in and the setting in which the trial was conducted. If the documents provide information on the context in which the study takes place, we will also include it.

Study Methods and Methodological Quality:
We will extract information about the randomization, quasi-experimental assignment, and other methodological aspects of the evaluation. The level of assignment and whether the study included multiple analyses at different levels will also be coded. Again, it is especially critical that information about the three key methodological issues [see section 2.3.1] in the implementation be extracted from each study report:

- Group equating procedure and issues
- Attrition
- Program fidelity issues

Intervention and Control Conditions data:
These items will solicit detailed descriptions of the intervention and control condition, including the "dosage" of the treatment being implemented, and the number of participants assigned to each group. We anticipate that the evaluations in this review sample will be comprised of a single intervention and a single control group. When this is not the case, we will select the most policy relevant groups to compute our experimental versus control condition contrast. In most cases, it will be the groups that experience the

greatest contrast between conditions, i.e., the most intensive intervention condition versus the least intensive control condition. We recognize the importance of documenting these decisions for full transparency.

The intervention will also be captured by a descriptive, open-ended item and as a categorical/nominal item, determined by the type of voucher the program employed. Likewise, the control condition will first be coded as a categorical/nominal variable (e.g., No Treatment; Treatment As Usual; Lesser Treatment) and then an open-ended item will provide more details and.

Participants in the Study:
These items solicit detail about the type of participants in the trials, including information on the country in which the study took place, the nationality of the participants, the age and school level targeted, gender, whether an urban or rural setting was involved or select areas within an urban ring (i.e. slums or barrios), and the socioeconomic status of the students.

2.3.2.3 Selection of outcome data for synthesis

Outcome data:
For each eligible study we will extract information on reported outcomes including impacts on access to education (private and public), learning, health, child labor, costs, and equity. We will also code any other outputs or data on key "mechanisms" that would provide clues as to why the intervention did or did not have its intended impact. Our primary focus will be on indicators of educational enrollment, attainment, and achievement (e.g., enrollment, matriculation, persistence, dropout, repetition, completion, test scores) for students considered "poor" either through a targeted program or implied by contextual information on locality of the program. We will also include, whenever possible, any "spillover" effects for public and non-voucher school in the same locality. However, the additional data on non-educational outcomes may serve as valuable descriptive or contextual support for the program's successes or shortcomings.

Handling multiple reports on the same experiment
Note that investigators may publish several articles on the same study. Our unit of analysis is the individual evaluation and not the individual research article, and so it is reasonable to extract information from all documents to complete the coding instrument for one experiment. When reports on the same study contain conflicting information, we will employ a number of strategies, including contacting the original investigator(s) for resolution.

2.3.2.4 Process used to combine/ synthesise data

Criteria for determination of independent findings
Each study will be represented by a *single effect size for each outcome variable, if possible,* to prevent the analysis from being compromised by non-independence (multiple effect sizes from one study). Our larger research indicates that most evaluations will use one follow-up measure. Thus, it is more than likely that we will employ a "First-Effects" analysis only.

For this review, we will keep outcomes distinct. That is, we will analyze enrollment, school attendance, dropout, equity, quality, and other learning or non-educational (health, behavioral, etc.) outcomes separately. We do not know as of yet how such outcomes will be reported, i.e., will they be prevalence measures (percentage of groups that enroll or attend) or incidence measures (the mean rate for some outcome of interest, such as the mean number of days attended per student). If results are varied and include prevalence and incidence rates, we will discuss the best way to report these (combine or separate out) and make such decisions explicit in our review.

However, as mentioned above, each study will contribute only one effect size to each outcome analysis. For example, a study may have reported data on enrollment, test scores and teacher attendance. If meta-analyses are conducted to synthesize outcomes in each of these three domains, the study would report one enrollment outcome to the enrollment meta-analysis and one outcome to the teacher attendance meta-analysis. In nearly all cases, this will be the first reported effect in the domain area (an outcome of enrolment at one year, for example, when we are looking at enrolment).

If regression-adjusted estimates are reported for the experimental versus control groups, we will rely on them for any quantitative synthesis since they theoretically reduce statistical "noise" that may have come from chance fluctuations or randomization violations (in the case of well implemented experiments) or uncontrolled variables (in the case of quasi-experiments). If there are multiple estimation models in the study, we will select the model that controls the most "noise", or includes the most covariates. Also note that we will standardize our variance estimates of the effect size in each individual study to take such covariates into account. While it is true that computing effect sizes from regression analyses is problematic, there are usually possibilities to use the binary proportion data to compute effect sizes. For example, most econometric studies report standardized coefficients representing the percentage gain in the outcome for treatment; with other data available in the report, this can be used to compute an effect size based on binary proportion differences between groups.

Some studies report analyses at multiple levels, i.e., for schools or localities and for studies. Our rule is to capture this information separately, but to compute effect sizes for the analysis done at the level of assignment. We will compute effect sizes based on both aggregate and individual N's. Based on our prior experience, most PIs report standardized effect sizes for test scores and binary proportion data for outcomes like enrollment and attendance. Effect sizes based on proportions do not change regardless of sample size, but the variances are much larger if aggregate units are used, due to the fact that sample population is so much smaller. However, PIs of primary literature in the developing world often report analyses based on individual N's even when aggregate units are assigned. Our approach is to compute effect sizes for both aggregate N's and individual N's, when possible to do so.

Some studies also report effects at all grade levels or by gender (e.g., Angrist et al., 2002). This is very important to policy and practice decision-makers. The main effect will again be computed at the larger analysis level, so that if schools are assigned to groups, the effect size will be computed for all schools in treatment versus all schools in control. However, we will record subgroup effects such as breakdowns by grade and gender.

Statistical procedures and conventions

The data will be entered into the Comprehensive Meta-Analysis (CMA), version 2. We will use CMA to statistically combine results from the evaluations.

We will use one outcome measure for this review. We will use a standardized measure of effect such as Cohen's effect size measure *(d), which will* be used to represent the standardized difference between experimental and control/comparison groups. Assuming the search yields a sufficient number of studies to carry out a quantitative analysis, each study will be code as a single case (row) in the database, with all quantifiable data extracted from each report.

Forest plots will be used to display the results from the effect sizes. The plot will display, for each study, the effect size, confidence intervals and significance level. The plot will also display the same for the average effect across studies. Because of the likely heterogeneity in interventions, samples, countries, and outcomes, we will assume random effects models in our analyses, which tends to be more conservative than the fixed effects approach. The random effects model assumes that effect sizes vary both as a function of

sampling error and genuine study level differences. In contrast to Analysis of Variance (ANOVA) (in which each "factor" can be used to specify fixed or random effects), the random effects model estimates the latter from the observed heterogeneity of effect sizes and uses this random effects variance component (tau-squared) to adjust the model, producing more credible standard errors.

For our analyses, we will conduct tests for heterogeneity (Q, tau^2, I^2) to determine if the average effect size is a good representation of the sample of studies being used in the analysis. We anticipate that heterogeneity will be present (assume we yield a sufficient sample of studies), given the variations in intervention type, nation, sample populations and the like in these development studies. However, moderator analyses have to be approached carefully, as they are often based on small numbers of studies (the "small cell" problem), and that such analyses can be significant by chance if large numbers of variables are considered (the "capitalizing on chance" problem). In the event that our searches yield a small number of high-quality studies, we may not be able to conduct any meta-analyses, let alone moderating analyses.

If our yield of studies is sufficient, however, the following candidates for moderating variables will be examined as a source of heterogeneity, by comparing the effect sizes (assuming a random effects model) for the following potential moderators:

Randomized versus non-randomized studies. An important question is whether the estimates from randomized experiments are different than those reported in studies that used non-random assignment. This review will examine the average effect size for experiments and compare it to the average effect size for quasi-experiments.

Different variations of school voucher programs. An important policy question is whether developing nations and donor agencies are getting more "bang for the buck" using one particular approach or another. We will use the categorical intervention variable for analysis and compare the average effect sizes for these groups.

World Bank Classification of Economies. Although it is not likely that the studies will report enough information to adequately assess the socioeconomic status of participants, we will examine findings based on the per capita income of the country.

School level. Interventions may be designed to increase access to private school at the primary or secondary level. We will examine effect sizes, if possible to do so, to determine the moderating influence of school level.

Gender. Programs in developing countries often target female enrollment as the primary goal in the program. Developing countries often have an education gap between male and female students. We will, if possible, examine the effect sizes based on what percentage of female participants are in the sample (e.g., 0-25%; 26-50%; 51-75%; 75-100%).

Publication bias. We do not anticipate, at this time, conducting a study of publication bias. In our larger review of the effects of school enrollment interventions in developing nations (Petrosino et al 2009), our searches have resulted in over 100 experimental and quasi-experimental studies. Many of these have been conducted by economists and our primary documents are reports (e.g., Angrist et al., 2001) that are technically not published but made available online through repositories such as the National Bureau of Economic Research (NBER) or Econpapers. However, we have found that most of these papers are eventually published within 1-3 years in economics journals. Therefore, papers may only be temporarily "unpublished." In this review, we will examine our final pool of "published" and "unpublished" studies and may revisit whether to conduct an analysis of publication bias.

Synthesis of descriptive data
We will synthesize descriptively the studies that do not meet the criteria for inclusion in effect size estimates. That is, we will number and describe the different types of interventions and the methods used to evaluate them, as well as the countries in which the interventions took place. We will discuss the findings of these studies in a narrative fashion, presenting the salient themes across the studies and providing specific examples from the reports.

2.4 Deriving conclusions and implications

Our review conclusions will be data-driven and based on the effect-size estimates in the review. We will draw conclusions, based on the pool of studies, about voucher programs' impact on access to a private school education for "poor" families in developing countries. The generalizability of the review must consider the studies used for the summary effects. If studies are isolated to one region, it may be inadvisable to interpret the results universally. We will rely on conversations among the authors and in consideration of what would strengthen the research base for a future version of the review draw out implications about research. We will examine what the implications of the findings might be for policy and practice, informed by our readings of the literature and conversations with staff from DFID, 3ie, and others.

References

Angrist, Joshua Eric Bettinger Erik Bloom Elizabeth King & Michael Kremer. Vouchers for private schooling in Columbia: Evidence from a randomized natural experiment. American Economic Review. 2002; 92(5): 1535-1558.

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, 847-862.

Arenas, Alberto. (2004). Privatization and vouchers in Colombia and Chile. International Review of Education. 50, 370-395.

Barrera-Osorio, F. (2007). The impact of private provision of public education: empirical evidence from Bogota's concession schools. Policy Research Working Paper Series 4121, The World Bank.

Behrman, J.R., Sengupta, P., & Todd, P.E. (2000). Impact of PROGRESSA on achievement test scores in the first year. International Food Policy Research Institute (IFPR).

Behrman, J.R., Parker, S.W., & Todd, P.E. (2010). *"Incentives for Students and Parents"*. A paper for conference on: "Educational Policy in Developing Countries: What Do We Know, and What Should We Do to Understand What We Don't Know?" (4-5 February 2011).

Birdsall, N. (1999). Comment: A vicious cycle. In V. Tanzi, K. Chu, & S. Gupta (Eds.), *Economic policy and inquiry*. Washington, D.C.: International Monetary Fund.

Birdsall, N., Levine, R., & Ibrahim, A. (2005). Toward universal primary education: Investments, incentives, and institutions. London: United Nations Development Programme.

Braun-Munzinger, Corinna. (2005). Education Vouchers – An International Comparison. Centre for Civil Society.

The Campbell Collaboration. (2011). "What is a systematic review?". Retrieved on 28 May 2011.
http://www.campbellcollaboration.org/what_is_a_systematic_review/index.php

Contreras, D. (2002) Vouchers, School Choice and Access to Higher Education, Yale University Economic Growth Center Discussion Paper No.845

Filmer, D., & Pritchett, L. (1998). The effect of household wealth on educational attainment around the world: Demographic and health survey evidence." Working Paper 1980. Washington, D.C.: The World Bank.

Filmer, D., & Pritchett, L. (2001). *Estimating Wealth Effects Without Expenditure Data— Or Tears: An Application to Educational Enrollments in States of India.* Demography, 38(1), 115-132.

Friedman, M. 1955. The Role of Government in Education. Economics and the Public Interest. Rutgers University Press.

Gauri, V. & Vawda, A. (2003). Vouchers for basic education in developing countries: A principal agent perspective. Policy Research Working Paper Series 3005. The World Bank.

Hanuum, E., & Buchmann, C. (2004). Global educational expansion and socioeconomic development: an assessment of findings from the social sciences. *World Development*, 33(3), 1-22.

Herz, B., & Sperling, G. (2003). What works in girls' education: Evidence and policies from the developing world. Washington, D.C.: Council on Foreign Relations.

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from www.cochrane-handbook.org.

Hsieh, C.T. & Urquiola, M. (2003). When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program. NBER Working Papers 10008, National Bureau of Economic Research, Inc.

Levin, Henry (2002). A Comprehensive Framework for Evaluating Educational Vouchers. Educational Evaluation and Policy Analysis, 24: 159–174.

Thapa, R. & Mahendra, S. (2011). *Subsidising education: are school vouchers the solution?*. International Initiative for Impact Evaluation. EQ Brief. July 2010. Retrieved April 22, 2011 www.3ieimpact.org/admin/pdfs/**Vouchers**_EQ_16_July_1.pdf

Kim, J., Alderman, H. & Orazem, P. (1998). Can private school subsidies increase schooling for the poor?: The Quetta Urban Fellowship Program. Policy Research Working Paper Series 11. The World Bank.

King, E., Rawlings, L., Gutierrez, M., Pardo, C., & Torres, C. (1997). Colombia's targeted education voucher program: Features, coverage, and participation. Working Paper No. 3, Series on Impact Evaluation of Educational Reforms, Development Economics Research Group, The World Bank.

McEwan, P.J. (2000). The potential impact of large-scale voucher programs. Occasional Paper. National Center for the Study of Privatization in Education. Teachers College, Colombia University, New York; p. 70.

McEwan, P.J., & Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. Educational Evaluation and Policy Analysis. 22(3), 213-239.

Ravallion, M. & Wodon, Q. (1999). *Does child labor displace schooling? - evidence on behavioral responses to an enrollment subsidy*. Policy Research Working Paper Series 2116, The World Bank.

Sapelli, C., & Vial, B. (2005) Private vs public school voucher schools in Chile: New evidence on efficiency and peer effects. Pontificia Unviersidad Catolica De Chile Institio De Economia, Working Paper 289.

Sapelli, C. (2010). Some thoughts on the evaluation of the Chilean voucher system. Journal of School Choice: Research Theory, and Reform. 4(2), 222-231.

Schady, N. & Araujo, M.C. (2006). Cash transfers, conditions, school enrollment, and child work : evidence from a randomized experiment in Ecuador. Policy Research Working Paper Series 3930, The World Bank.

Shafiq, N.M. (2010). Designing targeted educational voucher schemes for the poor in developing countries. International Review of Education. 56, 33-50.

West, E.G. (1997). Education vouchers in principle and practice: A survey. The World Bank Observer. 12(1), 83-103.

Yardley, J. (2005, March 13). China Plans to Cut School Fees For Its Poorest Rural Students. *New York Times*. p. 14. Retrieved from EBSCO*host*.

UNESCO (United Nations Educational, Scientific, and Cultural Organization) (2007). Education for all global monitoring report: Education for all by 2015: Will we make it? Paris: UNESCO.

The World Bank (2008). Tembon, M., & Fort, L. (Eds). *Girls' education in the 21ˢᵗ Century: Gender equality, empowerment and economic growth.* Washington, D.C.: The World Bank.

The World Bank (2009a). Six steps to abolishing primary school fees. Washington, D.C.: The World Bank.

The World Bank (2009b). Abolishing school fees in Africa: Lessons from Ethiopia, Ghana, Kenya, Malawi, and Mozambique. Washington, D.C.: The World Bank.

Appendices

Appendix 1.1: Authorship of this report

Authors

Trevor Fronius, Learning Innovations at WestE, Woburn, MA
Anthony Petrosino, Learning Innovations at WestEd, Woburn, MA
Claire Morgan, Learning Innovations at WestEd, Woburn, MA

Contact details

Trevor Fronius
Learning Innovations at WestEd
200 Unicorn Park Drive, 4th Floor
Woburn, MA 01801 USA
Tel: 781-481-1119
Fax: 781-481-1120
tfronius@wested.org

Acknowledgements

Appendix 2.1: Inclusion and exclusion criteria

(1) *The program evaluation have to be located in a country classified as a low-, lower middle-, or upper middle country* as defined by the World Bank at the time the intervention or program was implemented.

(2) *The evaluations must include outcomes that measure school vouchers* – as define in section *1.5 – impact on access to private schooling for poor people*. The outcomes will be identified based on the level of assignment for said intervention. We will collect all relevant outcome measures reported in the study, such as enrolment, student achievement and attainment outcomes, and school quality outcomes. Additionally, we will identify and summarize any cost-benefits information available in the report.

(3) *The evaluation study report is published or available between January 1991 and January 2011, without regard to language or publication type*. We will attempt to find English and non-English studies. In addition we will include both published and unpublished (e.g., dissertations, working papers, technical reports, conferences papers).

Our search will focus on identifying:

(4) *Randomized controlled trials (RCTs) or quasi-experimental evaluations with some evidence that the groups being compared are equivalent.* Our review will include evaluations that randomly assigned entities (at any level) to intervention or control conditions. In a previous review (Petrosino et al., 2009), the control or comparison group typically receives no intervention or continues to receive the same intervention as usual. Due to the fact that randomization is simply not possible for all studies (e.g., retrospective evaluations, post-hoc evaluations), we will target quasi-experimental designs that provide evidence that the comparison groups were equated (e.g., regression discontinuity design (RDD), propensity scores (PSM), covariate matching, and other pre or post intervention matching). Both types of experiments will be included in the summary analyses, and the study design will act as a moderator in our later analyses (*see section 2.3.2*).

Although RCTs and quasi-experimental designs with group equivalence will inform the effect size estimates, we will also include the following designs as part of the review:

(5) *Non-experimental and quasi-experimental without pretest group equivalency* (e.g., before and after studies) will be examined, but not included in effect size estimates.

(6) *High-quality descriptive quantitative and qualitative studies.* We will identify and examine non-causal descriptive studies, as long as there is a sufficient degree of transparency and detail as to be replicable, but we will not include these studies in the estimate effect sizes.

Appendix 2.2: Search strategy for electronic databases

The databases [Appendix 2.3] we will search tend to be unique in their query parameters. Thus, our search strategy will be conducive to the sensitivity of each database. We would prefer that our queries return many titles, rather than not capture relevant citations. All searches, where possible in the database, will search both titles and abstracts. In addition, where possible, we will search using thesaurus and other controlled vocabulary terms related to vouchers. We will search using terms for vouchers, such as "voucher", "subsidy", "subsidies", "subsidize", "subsidise", "school choice", "chit", "tax credit", "tax stipend", "stipend". In searches we have completed for previous reviews, and in the searches we have piloted for the present review, we have had most success in identifying relevant studies of interventions in developing countries by combining topic terms with the names of the developing countries of focus. We will also combine the voucher terms and country names with general terms for developing nations, such as "developing nation", "developing region" "developing country", "developing countries", "third world nation", "third world country", "third world countries", "third world region", "low income nation", "low income country", "low income countries", "low income region", "LIC country", "LIC countries", "emerging nations", "underdeveloped nations". Also, if the database is not focused on education, we will include in the string education terms such as "education" and "school", and if the database is focused on education, we will narrow the search to primary and secondary school.

As an example of our search strategy, this is a Boolean search of PAIS International, followed by a Boolean search of ERIC. These examples show searches using controlled terms. Such a search would be followed by a similar search of free-text terms searching all fields. Also, since ERIC limits the number of keywords that may be combined in each search, a few searches will be run in order to accommodate all of the country names.

EXACT("Education vouchers") OR EXACT("School choice") OR EXACT("Schools, Private: Tuition, fees, etc.") AND all((Afghanistan OR Albania OR "American Samoa" OR Angola OR Argentina OR Armenia OR Azerbaijan OR bangla desh OR "East Pakistan" OR "East Bengal" OR Belarus OR byelorussian OR Belize OR "British Honduras" OR Benin OR Dahomey OR Bhutan OR Bolivia OR "Bosnia Herzegovina" OR Botswana OR Bechuanaland OR Brazil OR Bulgaria OR "Burkina Faso" OR "Upper Volta" OR Burundi OR Cambodia OR "Khmer Republic" OR Kampuchea OR Cameroon OR "Cape Verde" OR "Central African Republic" OR Ubangi-Shari OR "Central African Empire" OR Chad OR Chile OR China OR Colombia OR Comoros OR "Democratic Republic Congo" OR Zaire OR "Belgian Congo" OR "Republic Congo" OR "Middle Congo" OR "Costa Rica" OR "Cote d'Ivoire" OR "Ivory Coast" OR Cuba OR Djibouti OR "French Somaliland" OR "Afars Issas" OR Dominica OR "Dominican Republic" OR Ecuador OR Egypt OR "Arab Republic Egypt" OR "El Salvador" OR Eritrea OR Ethiopia OR abyssinian OR Fiji OR Gabon OR "The Gambia" OR Georgia OR Ghana OR "Gold Coast" OR Grenada OR Guinea OR "French Guinea" OR Guinea-Bissau OR "Portuguese Guinea" OR Guyana OR "British Guiana" OR Haiti OR Honduras OR India OR Indonesia OR "Netherlands East Indies" OR Iran OR "Islamic Republic Iran" OR Persia OR Iraq OR Jamaica OR Jordan OR Kazakhstan OR Kenya OR Kiribati OR "Gilbert Islands" OR "Democratic Republic Korea" OR "North Korea" OR Kosovo OR "Kyrgyz Republic" OR Lao OR "Lao PDR" OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR "Macedonia FYR" OR Madagascar OR Malawi OR Nyasaland OR Malaysia OR Malaya OR Maldives OR Mali OR "French Sudan" OR "Marshall Islands" OR Mauritania OR Mauritius OR Mayotte OR Mexico OR Micronesia OR "Federal States Micronesia" OR Moldova OR moldavian OR Mongolia OR Montenegro OR "Federal Republic Yugoslavia" OR Morocco OR Mozambique OR Myanmar OR Burma OR Namibia OR "South-West Africa" OR Nepal OR Nicaragua OR Niger OR Nigeria OR Pakistan OR Palau OR Panama OR "Papua New Guinea" OR Paraguay OR Peru OR Philippines OR Poland OR Romania OR "Russian Federation" OR Rwanda OR Samoa OR "Western Samoa" OR "Sao Tome Principe" OR Senegal OR Serbia OR Seychelles OR "Sierra Leone" OR "Solomon Islands" OR Somalia

OR "South Africa" OR "Sri Lanka" OR Ceylon OR "St. Kitts" OR Nevis OR "St. Lucia" OR "St. Vincent" OR Grenadines OR Sudan OR "Anglo-Egyptian Sudan" OR Suriname OR "Dutch Guiana" OR Swaziland OR "Syrian Arab Republic" OR Syria OR Tajikistan OR "Tajik Soviet Socialist Republic" OR Tanzania OR Thailand OR Siam OR timor leste OR "East Timor" OR "Portuguese Timor" OR Togo OR "French Togoland" OR Tonga OR Tunisia OR Turkey OR Turkmenistan OR Uganda OR Ukraine OR Uruguay OR Uzbekistan OR "Uzbek Soviet Socialist Republic" OR Vanuatu OR "New Hebrides" OR Venezuela OR Vietnam OR "West Bank and Gaza" OR "West Bank" OR Gaza OR Yemen OR Zambia OR Zimbabwe OR "developing nation" OR "developing region" OR "developing country" OR "third world nation" OR "third world country" OR "third world region" OR "low income nation" OR "low income country" OR "low income region" OR "upper middle income nation" OR "upper middle income country" OR "upper middle income region" OR "lower middle income nation" OR "lower middle income country" OR "lower middle income region" OR "impoverished country" OR "impoverished region"))

(Thesaurus Descriptors:"Educational Vouchers" OR Thesaurus Descriptors:"school vouchers" or Thesaurus Descriptors:"tuition tax credits" or Thesaurus Descriptors:"tuition grants" or Thesaurus Descriptors:"private school aid" or Thesaurus Descriptors:"school choice" or Thesaurus Descriptors:"fellowships") and (Keywords: "African Republic" or Keywords: Ubangi-Shari or Keywords: "Central African Empire" or Keywords: Chad or Keywords: Chile or Keywords: China or Keywords Colombia or Keywords Comoros or Keywords: Congo or Keywords: Zaire or Keywords: "Costa Rica" or Keywords: "Cote d'Ivoire" or Keywords: "Ivory Coast" or Keywords: Cuba or Keywords: Djibouti or Keywords: "French Somaliland" or Keywords: "Afars and Issas" or Keywords: Dominica or Keywords: "Dominican Republic" or Keywords: Ecuador or Keywords: Egypt or Keywords "Arab Republic Egypt" or Keywords: "El Salvador" or Keywords: Eritrea or Keywords: Ethiopia or Keywords: Abyssinia or Keywords: Fiji or Keywords: Gabon or Keywords: "The Gambia" or Keywords Georgia or Keywords: Ghana) or Thesaurus Descriptors:"developing nations" or Thesaurus Descriptors: " and (Education Level:"Elementary Education" OR Education Level:"Elementary Secondary Education" OR Education Level:"Grade 1" OR Education Level:"Grade 2" OR Education Level:"Grade 3" OR Education Level:"Grade 4" OR Education Level:"Grade 5" OR Education Level:"Grade 6" OR Education Level:"Grade 7" OR Education Level:"Grade 8" OR Education Level:"Grade 9" OR Education Level:"Grade 10" OR Education Level:"Grade 11" OR Education Level:"Grade 12" OR Education Level:"High Schools" OR Education Level:"Intermediate Grades" OR Education Level:"Junior High Schools" OR Education Level:"Kindergarten" OR Education Level:"Middle Schools" OR Education Level:"Primary Education" OR Education Level:"Secondary Education")Publication Date:1991-2011

The appendices to our final review report will carefully document all keywords used for each database to permit replication.

Appendix 2.3 Glossary of abbreviations

Abbreviation	Term
RCT	Randomized controlled trials
QED	Quasi-experimental design
NGO	Non-governmental organization
CCT	Conditional cash transfer
SES	Socioeconomic status
RDD	Regression discontinuity design
PSM	Propensity score matching
IV	Instrumental variable
DID	Difference in Differences
CMA	Comprehensive Meta-Analysis, version 2
ANOVA	Analysis of Variance

Appendix 2.4: List of bibliographic databases to be searched

- Australian Education Index
- British Education Index
- British Library
- CBCA Education
- EBSCO Megafile (i.e., Academic Search Premier, Education Full Text (H.W. Wilson), Education Index Retrospective: 1929-1983 (H.W. Wilson), ERIC, Family & Society Studies Worldwide, MasterFILE Premier, MLA Directory of Periodicals, MLA International Bibliography, Race Relations Abstracts, and Urban Studies Abstracts)
- Econlit
- Eldis
- Google and Google Scholar
- IDEAS-RESEARCH PAPERS IN ECONOMICS http://ideas.repec.org/
- Index to Current Urban Documents
- Inside Info Plus (British Public Library)
- International Bibliography of Social Science
- ISI Web of Science (Social Sciences Citation Index)
- JOLIS (World Bank, International Monetary Fund, International Finance
-     Corporation)
- PAIS Archive
- PAIS International
- Periodical Contents Index
- PolicyFile
- Proquest Dissertations and Theses
- Sage Journals Political Science Full-Text Collection
- Sage Journals Sociology Full-Text Collection
- Sage Urban Studies Abstracts
- Social Science Research Network
- Sociological Abstracts
- UNESCO (UNESDO and UNESBIB)
- Worldwide Political Science Abstracts

Appendix 2.5: Journals to be handsearched

Because electronic searches often miss relevant studies, we will hand search the table of contents, and the abstracts when necessary, of all issues of the journals most likely to publish studies on this topic. From our initial searches, the five journals that we have identified for hand search are: *Economic Development and Cultural Change, International Journal of Educational Development*, *Journal of Development Economics*, *World Bank Research Observer*, and the *World Bank Economic Review*.

Appendix 2.6: Draft coding tool

DFID Review: Impact of School Vouchers on Access to Private Education in Developing Countries
Impact Evaluations CODING INSTRUMENT

Coder:
- ❑ Trevor Fronius
- ❑ Anthony Petrosino
- ❑ Claire Morgan
- ❑ Other _____

I. RESEARCHER AND STUDY CHARACTERISTICS

What year was the primary document published? _____

What was the type of document?
- o Book
- o Book Chapter
- o Government Report
- o Technical Report (reports by non-Govt. research firms, e.g. Mathematica)
- o NGO Report (e.g., World Bank, Poverty Action Lab)
- o Journal (peer reviewed)
- o Dissertation
- o Conference Paper
- o Other

In what country did the evaluation take place? _____

World Bank country classification at time of study
- ❑ Lower Income
- ❑ Lower Middle Income
- ❑ Upper Middle Income

What was the setting for the evaluation? _____
(e.g. urban/rural; slum/non-slum, etc.)

Who conducted the evaluation? (e.g., medical researchers, economists, etc. May be an assumption based on the affiliation)
_____

Baseline enrolment data: Males _____ Females _____
(Use enrolment rates as close in proximity to intervention setting as possible, but if only national rates available, use those)

II. STUDY METHODS AND METHODOLOGICAL QUALITY

What method of assignment was used to assign or form groups?

- ❑ Random Assignment
- ❑ Non-Random Assignment (Quasi-experimental)
- ❑ Combination of Random Assignment and Non-Random Assignment (e.g., randomization only after oversubscription of available "spots")

If non-random assignment, what procedure was used to assign or form groups?

- ❑ Regression Discontinuity Design
- ❑ Statistical Matching
- ❑ No Equating of Groups
- ❑ Other (Indicate: _____)

If statistical matching used, what procedure was used to match?

- ❑ Propensity Scores
- ❑ Covariate matching
- ❑ Other (Indicate: _____)

At what level was assignment made?

- ❑ Village/Neighbourhood
- ❑ School
- ❑ Classroom
- ❑ Household
- ❑ Individual
- ❑ Other (Indicate:_____)

Methodological Threats to Evaluation Design

Threat	Did it exist?	How extensive? (Percentage of sample)	What did authors do to address?	Rate the Threat to Evaluation Findings about Enrolment (None/Low/Moderate/High)
Crossovers	YES/NO			
Attrition from Original Study Sample	YES/NO			
Attrition of Students from Larger Aggregate Unit Assignment	YES/NO			
Differential Attrition	YES/NO	(Percentage difference between groups)		
GROUP INEQUITY AT PRETEST				

Number of variables examined	Number of statistically significant differences	What did authors do to address?	Rate the threat to evaluation findings about enrolment (None/Low/Moderate/High)

What was the overall methodological quality?
- ❑ High
- ❑ Medium
- ❑ Low

III. INTERVENTION AND CONTROL CONDITIONS

What was the type of intervention?

- ❑ Unrestricted, Flat Voucher system
- ❑ Targeted Voucher system
- ❑ Means-tested Voucher system
- ❑ "Follow the Child" system
- ❑ Other (Indicate: _____)

Describe the intervention group below, with particular attention to the "dosage" of the treatment:

_____
_____
_____
_____
_____

How many cases were randomized or assigned to this group? _____

Program Implementation/Fidelity

Program Implementation Issues Mentioned by Authors (Not Possible but Actual)	What did authors do to address?	Rating

Please provide simple program theory (or mechanisms for why the intervention should work):

_____
_____
_____
_____

What is the control or comparison condition?

- ❑   No Treatment Group
- ❑   Wait-List Control
- ❑   Treatment as Usual Group
- ❑   Placebo
- ❑   Lesser dose of the same treatment
- ❑   Entirely different treatment than what Experimental got
- ❑   Other (Indicate:_____)

Describe the control or comparison condition (including "dosage" and where it came from if applicable):

_____
_____
_____
_____
_____

How many cases were randomized or assigned to this group? _____

IV. PARTICIPANTS IN THE STUDY

Type of school(s):
- ❑   Public
- ❑   Private
- ❑   Religious
- ❑   NGO

Age/grade   _____

School Level(s)
- ❑   **Primary**
- ❑   **Secondary**
- ❑   **Unknown**

Percentage of participants that were female   _____

Poverty/SES _____

V. OUTCOMES

What was the overall conclusion or investigator-reported result (IRR)?

- ❑   Positive
- ❑   Null or no effect
- ❑   Negative effect
- ❑   Mixed-Can't discern

SIMPLY INDICATE THE EDUCATION AND NON-EDUCATION OUTCOMES AND WHEN REPORTED (TIME INTERVAL)

Education/Learning Outcome	Outcome Measurement at What Time Intervals (only those in which data points are reported, e.g., 6 months, 12 months, etc.)	NON-EDUCATION OUTCOMES	Outcome Measurement at What Time Intervals (only those in which data points are reported, e.g., 6 months, 12 months, etc.)
Enrolment			
Attendance			
Dropout			
Test Scores			
Grades			
School Quality			
Equity			
Other (List each in a new row)			

Were subgroup effects for treatment reported? (Yes/No)

     If so: List:

_____

_____

_____

—

Was any cost-benefit or economic analysis reported? (Yes/No)

Indicate outcome of economic analysis:

- ❑ Program Group is more efficient option
- ❑ Comparison/Control Group is more efficient option
- ❑ Program Group is more efficient than policy alternatives
- ❑ Policy Alternatives are more efficient than program group
- ❑ No clear distinction between the two groups

ANY OTHER COMMENTS ON THE PROGRAM OR EVALUATION

The **Evidence for Policy and Practice Information and Co-ordinating Centre** (EPPI-Centre) is part of the Social Science Research Unit (SSRU), Institute of Education, University of London.

The EPPI-Centre was established in 1993 to address the need for a systematic approach to the organisation and review of evidence-based work on social interventions. The work and publications of the Centre engage health and education policy makers, practitioners and service users in discussions about how researchers can make their work more relevant and how to use research findings.

Founded in 1990, the Social Science Research Unit (SSRU) is based at the Institute of Education, University of London. Our mission is to engage in and otherwise promote rigorous, ethical and participative social research as well as to support evidence-informed public policy and practice across a range of domains including education, health and welfare, guided by a concern for human rights, social justice and the development of human potential.

This document is available in a range of accessible formats including large print. Please contact the Institute of Education for assistance:

telephone: +44 (0)20 7947 9556   email: info@ioe.ac.uk