



NOTES FOR REVIEW GROUPS

If you are familiar with using ‘styles’ in Word, please format all the text using only styles beginning with ‘EPPI’.

If you are not familiar with using ‘styles’ in Word, please format your document in the simplest way possible, while making sure that it is clear where headings fit within the hierarchy of the document.

As space is automatically left between each line, so you do not need to add in extra line breaks between each paragraph.

Please divide up chapters with Page Breaks, not Section Breaks. Only use Section Breaks if you need to alternate between portrait and landscape pages.

Structure for a protocol

Main title	A systematic review of the evidence of the impact in the medium-longer term of the abolition of school fees in low-income developing countries.
Sub title	
Review group	WestEd
Section	PROCOTOL
Authors IN ORDER OF CREDIT (Please include first and surnames, institutions. Include titles - Dr, Prof - if you want them to be used.)	Claire Morgan, WestEd Anthony Petrosino, WestEd Trevor Fronius, WestEd
EPPI-Centre reference number	[To be completed by EPPI-Centre]
Month/year of publication	[To be completed by EPPI-Centre]
This report should be cited as...	
Contact details (address, phone number, email)	Claire Morgan Learning Innovations at WestEd 200 Unicorn Park Drive, 4th Floor Woburn, Massachusetts 01801 Tel: 781-481-1119 Fax: 781-481-1120 cmorgan@wested.org

Institutional base	WestEd
Review Group (with institutions)	
Advisory group (with institutions)	
Conflicts of interest (if any)	None
Acknowledgements	We acknowledge support from the UK Department for International Development (DFID) and the International Initiative for Impact Evaluation (3ie).

1. Background

1.1 Aims and rationale for review

It has been well established that education is key to economic development and social welfare. Investments in education 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 contributes to increased political participation and more equitable sharing of economic and political power (Birdsall, 1999). Education for girls is particularly critical, as 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, 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 can be attributed in part to the private costs associated with sending children to public school (e.g. Filmer & Pritchett, 1998; Bentaouet-Kattan and Burnett, 2004). These household costs include lost work opportunities for children who attend school as well as direct costs in the form of school fees. Direct school user fees paid by households are common in developing countries and represent a percentage of all primary education costs ranging from 8% in Indonesia to 80% in Cambodia (Bentaouet-Kattan and Burnett, 2004). Eighty-three percent (83%) of World Bank client countries surveyed in 2005 assessed user fees (Bentaouet-Kattan, 2006). These costs, which include fees for books and uniforms, community and PTA contributions, exam fees, and, less commonly, tuition represent a large percentage of total household spending and are particularly burdensome for those families that face tough choices about which children to send to school and for how long (World Bank, 2009a). School fees can be seen as a form of regressive taxation that disproportionately burdens the poor, including the most vulnerable, such as girls, child laborers, and children living in remote areas (Bentaouet-Kattan, 2006; World Bank, 2009a).

During the independence movements of the 1960's, many developing nations made investments in free basic education in order to build capacity for more equitable participation in economic growth and political participation, but these policies were abandoned during the 1980's, as governments found it increasingly difficult to provide for expanding school systems during times of low economic growth (World Bank, 2009b). More recently, motivation has been building to abolish user fees for basic education in developing nations, spurred by initiatives such as Education for All, the Millennium Development Goals, and the School Fee Abolition Initiative launched by UNICEF and The World Bank in 2005.

Several countries have experienced large increases in enrolment after abolishing school fees. For example, in Kenya, primary school enrolments increased from 5.9 million in 2002 to 7.12 million in 2004; Timor Leste saw a 9.5 percent increase between 1999 and 2001 (Bentaouet-Kattan, 2006 in World Bank, 2009a). In the year following fee abolition, Malawi and Uganda experienced enrollment increases of 51 and 68 percent respectively (World Bank, 2009b). While proponents claim that the abolition of user fees can reduce disparities and improve educational equity (Bentaouet-Kattan, 2006), dramatic increases in enrolment may be accompanied by tradeoffs in educational quality (World Bank, 2009b). In addition, abolishing school fees alone is not 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, 2009a).

We believe that a systematic review that focuses specifically on the evidence of the impact of eliminating school fees 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. We will systematically identify and synthesize studies of interventions in developing countries that evaluate the elimination of school user fees paid by households—including the five fee categories

identified by the World Bank (tuition, uniforms, textbooks, PTA contributions, other materials/activities). These interventions may be carried out at the national level and provide for universal free primary education, or may be carried out by NGOs targeting specific regions, schools, or vulnerable groups. These interventions include programs that provide fellowships or scholarships provided they explicitly are intended to eliminate user fees for households.¹

1.2 Definitional and conceptual issues

The review takes its conceptual framework from demand-side economics of education financing, which establishes the negative price elasticity of demand (other things being equal) for children's education, particularly that of girls, in poor households in low-income countries (e.g. Birdsall & Orivel, 1996; Gertler & Glewwe, 1989; in Hillman & Jenker, 2002). While school fees can remove supply-side limitations, they exclude those households most unwilling or unable to pay the fees, due to demand-side constraints such as opportunity costs of lost child labor, household contributions of children, low expectations of returns to education, unavailable credit markets, and social norms that discourage school participation (Hillman & Jenker, 2002). We will identify and code studies that test the hypothesis that eliminating user fees increases demand for education, as demonstrated by increased enrollment and persistence and decreased dropout. However, the relationship between school fees and demand could also be positive if increased school quality overcomes opportunity cost impediments (Hillman & Jenker, 2002). In addition to enrollment and persistence outcomes, we will also code student achievement and other school quality outcomes, such as pupil-teacher ratio, where available.

1.3 Policy and practice background

School fee abolition policies are supported by many development organizations and aid agencies and have been spurred worldwide by initiatives such as Education for All, the Millennium Development Goals, and the School Fee Abolition Initiative. Promises to abolish school fees are often politically motivated and featured in election campaigns, or may be part of a wider policy reform, often including educational decentralization. Primary school fees are most commonly targeted for elimination, as labour market returns to primary schooling in the developing world are generally greater for primary school than for secondary, and the costs of primary schooling are lower. School fee abolition policies may or may not include private schools. School fees may be abolished through a "big bang" approach, officially eliminating universal free primary education (although school fees often continue to be assessed at the school/community level), or through a phasing-in approach by region or grade/age level or through targeted exemptions aimed at vulnerable groups.

Research indicates that little of the central government education budget actually gets to the schools in some developing nations, sometimes due to corruption (Reinikka and Swenson, 2007). Thus, school fees charged at the local level are needed to supplement both teachers' salaries and to purchase school equipment (Grogan, 2006). One concern expressed in nations that have eliminated school fees is whether the resources will be available to support teacher payroll and other needs.

Countries that have eliminated school fees have, however, witnessed large increases in enrolment. For example, Uganda, using a "big bang" approach, experienced a 68 percent increase in gross enrolment the year following school fee abolition; Tanzania saw a 33% increase using a phased-in approach (Bentaouet-Kattan, 2006). In addition, access to education by the poor and by other vulnerable groups, such as girls and orphans, has

¹ Conditional cash transfers (CCTs)—payments made to households conditional on students being enrolled in school—represent an intervention with goals and outcomes that overlap with school fee elimination. Comparing the effectiveness of CCTs to school fee elimination, while beyond the scope of the current review, is an interesting avenue for further research which will be partially explored through a 3ie systematic review we are currently conducting examining school enrollment interventions in developing nations.

increased in countries that have eliminated school fees. However, such rapid surges in enrolment can result in decreased educational quality. The extent to which this is the case may depend upon the extent of planning prior to fee elimination and the strategies put in place to cope with the surge in enrolment. For example, Ghana used a combination of measures, including training additional teachers and ensuring the provision of additional textbooks, and saw marked improvement in test scores. In Malawi, on the other hand, massive growth in enrolment outstripped resources and led to a serious decline in school quality (World Bank, 2009b). Based on the countries' experiences with school fee abolition, the School Fee Abolition Initiative (SFAI), launched in 2005 by UNICEF and The World Bank (2009a), provides six steps for the successful abolition of school fees, including:

- Ensuring leadership and participation from the start;
- Assessing the situation;
- Establishing priorities;
- Estimating costs and identifying financing;
- Planning for quality; and
- Strengthening school governance and accountability.

1.4 Research background

The literature firmly establishes that user fees—which often constitute a large portion of overall education expenditures—inhibit school enrollment and persistence in developing countries (Bentaouet-Kattan, 2006). However, precise information on the types and extent of user fees, which are often unofficially assessed and sporadically reported, is difficult to obtain. In 2001 and 2005, the World Bank conducted surveys of World Bank education task team leaders for 93 countries regarding the prevalence of school fees in their countries and found that only 16 of the countries surveyed have no fees (Bentaouet-Kattan and Burnett, 2004; Bentaouet-Kattan, 2006). The survey further found that five main types of fees are collected, including tuition fees, textbook charges, uniform fees, PTA/community contributions, and other school-based activity fees. Of these fees, the most commonly collected were PTA/community contributions, which are also the most difficult to track, as they are often collected unofficially/illegally (Bentaouet-Kattan, 2006). In addition, the survey found school fees to be a significant barrier to the transition from primary to secondary school, as secondary school fees are more prevalent and substantial than those assessed at the primary levels. However, the survey also identified a growing trend toward the elimination of school fees. For example, between 1994 and 1999, only three countries had implemented this policy, while between 2001 and 2005, 13 countries abolished school fees (at least in official policy) (Bentaouet-Kattan, 2006).

Although surges in enrolment following the abolition of school fees in many developing countries have been demonstrated, as described above, relying on gross or net enrollment data provides a very limited picture of the impact of this policy. However, there is little empirical evidence available to date on the impact of eliminating school fees on educational outcomes in developing nations (Grogan, 2006). Using national survey data and employing regression discontinuity difference-in-difference estimation techniques, Grogan (2006) conducted the first quantitative analysis of the effects of introducing universal primary education policy in an African country. The study found that the elimination of school fees increased the probability that an individual attended school before age 8 by about 10%, with similar findings for both boys and girls. However, the policy was also found to cause a 10% decrease in the probability that a publicly schooled child of a given age and socioeconomic characteristics was able to complete a simple reading test.

To our knowledge, a systematic review of similar studies evaluating the impact of school fee abolition policies (including interventions such as establishing free universal primary education, eliminating textbook or uniform fees, or targeted fee exemptions for marginalized groups) has not been conducted. Through this review, extensive searching, including hand searches, examining grey literature, and contacting experts in the field, will identify rigorous experimental studies, as well as high quality non-experimental causal

evaluations conducted on this topic, to determine the size and quality of the evidence base. In addition, while surges in enrollment have been documented immediately following school fee abolition initiatives, it will be important to establish whether studies have been conducted of the impact of abolishing user fees for education in the medium-longer term, such as whether enrollment and achievement gains are sustained and whether overcrowding and other school quality issues are evident.

Although meta-analysis will not be possible if the yield of experiments or quasi-experiments proves to be very low, there is value in establishing the dearth of rigorous studies in an area currently receiving considerable support by development agencies, NGOs, and academic institutions. In the medical field, Cochrane reviews (about 10%) sometimes report a yield of zero studies. The importance for medical intervention decisions of documenting the areas in which new clinical trials need to be funded and implemented is similar for education development interventions.

1.5 Objectives

For this project, we will be collecting studies that respond to the question: What is the evidence of the impact in the medium-longer term of the abolition of school fees in low-income developing countries? We will assess the breadth and quality of the literature base and synthesize the available empirical evidence of the effect of eliminating primary and secondary public or private school fees [including tuition fees, fees for textbooks and other instructional materials, compulsory uniforms, Parent Teacher Association (PTA) dues, community contributions to district education boards, and other school-based activity fees (exam fees, etc.) in lower income countries, as defined by the World Bank--at the universal, community/district/targeted group, or the grade/age group or school level--on all outcomes reported, such as primary and secondary school enrollment; gender parity in enrollment; student persistence, completion, repetition, and dropout; student achievement; and school/educational quality indicators, such as pupil-teacher ratio, pupil-textbook ratio, per-pupil expenditure, infrastructural adequacy, and other indicators of overcrowding or declining educational quality.

2. Methods used in the review

2.2 User involvement

2.2.1 Approach and rationale

In order to reach potential stakeholders and users of the review, we will present the findings of the review 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 also pursue publication of the review in a peer-reviewed journal. In addition, we will share the review with organizations and groups that are currently driving policy in this area, including the World Bank, UNESCO, UNICEF, and SFAI. Also, throughout the process of identifying eligible studies for the review, we will make contacts with some of the authors in the field to request their help in identifying relevant studies. We will share the completed review with these authors as an additional dissemination effort. We will also explore ways to share the review with policymakers in the countries for which eligible studies were identified. We are committed to working closely with DFID to disseminate the results in other ways, including any organized meetings and specialized publications.

2.2 Identifying and describing studies

2.2.1 Defining relevant studies: inclusion and exclusion criteria

For this project, we will only include those studies that have the following characteristics:

(1) *The evaluations have to have taken place in a country classified as a low-income economy by the World Bank at the time the intervention being studied was implemented.*

(2) *The evaluations have to assess the impact of eliminating primary or secondary public or private school fees* [including, tuition fees, fees for textbooks and other instructional materials, compulsory uniforms, Parent Teacher Association (PTA) dues, community contributions to district education boards, and other school-based activity fees (exam fees, etc.)] whether at the universal level or the community/group, district, school, or grade/age level. We will collect all outcome measures reported in the evaluations, such as impact on enrolment, persistence, achievement, and school quality, and any cost-benefit information provided.

(3) *The evaluation study report is published or available through December 2009, without regard to language or publication type.* We will attempt to find English and non-English studies. In addition, we will include published and unpublished studies (e.g., from conference papers, dissertations, technical reports).

Our search will focus on identifying:

(4) *Randomized controlled trials or quasi-experimental evaluations with some evidence that the groups being compared are equivalent.* Our review includes evaluations that randomly assign entities (at any level) to intervention or control conditions. However, because randomization is not possible in certain evaluation scenarios (e.g., in retrospective evaluations or when the program has already been implemented), we will also include evaluative studies that use quasi-experimental designs, provided that they offer evidence that the comparison groups were equated. Such quasi-experiments include regression discontinuity design, propensity scores, covariate matching, and other pre or post intervention matching. We include both types of studies and will include study design as a moderator in our later analyses (see discussion on moderator variable analysis in later section).

In addition, we will include:

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

(6) *High-quality descriptive quantitative studies.* We will identify and examine non-causal descriptive quantitative studies, as long as they include sufficient methodological detail as to be replicable, but will not include effect sizes for these studies.

(7) *Qualitative studies.* We will also examine qualitative studies that shed light on possible theory, implementation and context issues, but will not include them in effect size estimates.

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, the University of Pennsylvania, Boston Public Library, and the University of Massachusetts, including ERIC, British Education Index, PAIS International/Archive, Sociological Abstracts and World Bank Documents. 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.

(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.

(4) *Citation chasing.* The reference section of every retrieved report will also be checked to determine whether any possible eligible evaluations are listed. 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, and Chinese who can translate abstracts or full text documents in non-English to determine their eligibility for this review.

(5) *Contacting the "informal college" 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 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, including a preliminary listing of eligible studies that we have identified, so that they can alert us to any missed studies. A full listing of all persons contacted will be included in the final review.

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. We will review all citations and determine if the cited study should proceed to a second screening, 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 eliminates primary or secondary school fees. Studies proceeding to this second stage will be reviewed by two authors, and disagreements will be resolved by consensus among all three authors.

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

2.2.4 Characterising included studies

Included studies will be identified by overall methodology (e.g. RCT, QED, descriptive) and by specific methodology (e.g. regression discontinuity, difference-in-difference, benefit incidence). In addition, studies will be characterized by specific type of fee abolition intervention (e.g. elimination of all fees, elimination of textbook fees, elimination of uniform fees) and by group or level targeted (e.g. universal free primary education, fee elimination for certain age/grade levels, targeted exemptions for vulnerable groups).

2.2.5 Identifying and describing studies: quality assurance process

To ensure that we achieve good coding reliability, we will have two of the coauthors read and record information from all reports. 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

2.3.1 Assessing quality of studies

For the evaluative studies that are the focus of the review and for which effect sizes will be reported, 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. It is especially critical that information about three key issues in the implementation be extracted from each study report to determine study quality:

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.

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 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 capture any data, qualitatively, on the types of program withdrawals in our comments section.

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."

For the non-experimental and quasi-experiments without pretest group equivalency and for the non-causal quantitative and qualitative studies, we will only include studies in which the methodology is sufficiently rigorous in terms of systematically and transparently collecting and analyzing the data to make clear conceptual links between the data and the conclusions drawn.

2.3.2 Overall approach to and process of synthesis

Extracting Information from Each Study

We have designed a preliminary instrument to guide us in recording information from each study (see Appendix 2.4). Although the instrument contains several open-ended items, these will be collapsed when appropriate into a smaller number of categories to permit further analysis.

The instrument has items 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. It is especially critical that information about three key issues in the implementation be extracted from each study report:

- How the groups were equated and whether any problems with equating were reported (see above description).
- 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 (see above description).
- Whether the program experienced significant implementation and fidelity problems (see above description).

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.

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, and the socioeconomic status of the students.

Outcome data:

For each eligible study we will extract information on reported outcomes including impacts on access to education, 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.

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.

Criteria for determination of independent findings

Each study will be represented by a *single effect size* to prevent the analysis from being compromised by non-independence (multiple effect sizes from one study). Although some evaluations may report just a single outcome at one time interval, it is more likely that evaluation reports will include analyses at various time intervals and may use various constructs that reflect school enrollment and persistence. Therefore, decisions have to be made about what outcome will represent the effect size for that study.

	Enrollment	Attendance	Dropout	Test Scores	Grades
First Effect					
Middle Effect					
Longest Effect					

For this review, we will keep outcomes distinct. That is, we will analyze enrollment, school attendance, dropout, and other learning or non-educational (health, behavioral, etc.) outcomes separately. The Table provides an example of how five such outcomes would be captured, but there could be others (e.g., matriculation). 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.

We also propose to report three different analyses to handle the studies that report outcome data at various time intervals. As the Table indicates, we will report effect sizes at first follow-up (the first time interval reported), the middle effect (the middle time interval closest to the exact point between the first and longest), and the longest effect (the effect size for the longest follow-up period). If one time interval (e.g., 1 year) is reported in the study, it will be used in all three analyses. If two time intervals (e.g., 6 months, 1 year) are reported, the results will be averaged and the mean will be reported for the "middle effect." If more than three time intervals are reported (e.g., 6, 12, 18 and 24 months), we will select the result that is closest to the exact middle. In this instance, the exact middle between 0-24 months is 12 months and would be reported as the middle.

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).

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 code information, however, about the analyses done at the student, family or school level.

Some studies also report effects at all grade levels. 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*), *that will* be used to represent the standardized difference between experimental and control/comparison groups. For the

quantitative meta-analysis, if possible to conduct, each study will be a single case (row) in the database, and data on over 100 variables will be extracted from study reports.

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. Note that this will be reported assuming a random effects model, and the estimate will be weighted by sample size. When describing results in the text, we will report the effect size, the confidence intervals and whether the analysis indicates that the result is statistically significant.

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.

For our analyses, we will conduct tests for heterogeneity 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, given the variations in intervention type, nation, sample populations and the like in these development studies. Using CMA, we will confirm heterogeneity in each summary analysis (of each outcome at each of the three time intervals: first, middle and longest) through the Q-Value, which is reported as a summary indicator of the extent of variance across studies in the sample.

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). We anticipate finding a very small number of studies, and therefore will not be able to conduct 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 elimination fee policies. 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 recode the detailed intervention categories into discrete groups for analysis and compare the average effect sizes for these groups.

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

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 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. We also do not anticipate finding many studies before 2001, but if that assumption does not hold true, we can also consider examining the year of publication as a moderating variable.

Treatment of descriptive research

Descriptive quantitative 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. It is very unlikely, because of the topic area and our focus on experiments and quasi-experiments that we will uncover much qualitative

research. However, we will code the presence or absence of ancillary qualitative studies, what the studies focused on, and what the main findings are. Certainly, 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.

2.3.2.1 Selection of studies for synthesis (if not all studies that are included in the synthesis)

See above.

2.3.2.2 Selection of outcome data for synthesis

See above

2.3.2.3 Process used to combine/ synthesise data

See above

2.4 Deriving conclusions and implications

Our review conclusions will be data-driven and based on the results in the review. We will draw conclusions about abolishing different types of school fees based on the effects reported across included evaluations. We will draw out implications about research based on conversations among the authors and in consideration of what would strengthen the research base for a future version of the review. 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

- Bentaouet-Kattan, R. (2006). Implementation of free basic education policy. Washington, D.C.: The World Bank.
- Bentaouet-Kattan, R., and Burnett, N. (2004). User fees in primary education. Washington, D.C.: The World Bank.
- 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.
- Evans, D., Kremer, M., & Ngatia, M. (2008). The impact of distributing school uniforms on children's education in Kenya. World Bank Working Paper.
- Filmer, D., and 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.
- Glewwe, P., Kremer, M., & Moulin, S. (2000). Textbooks and test scores: Evidence from a prospective evaluation in Kenya. World Bank Working Paper.
- Grogan, L. (2006). Who benefits from Universal Primary Education in Uganda? Department of Economics, University of Guelph, ON.
- Hanuam, 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.
- Hillman, A., & Jenkner, E. (2002). User payments for basic education in low income countries. IMF Working Paper WP/02/182.
- Kremer, M., Miguel, E., & Thornton, R. (2005). Incentives to learn. *Education Next: a Journal of Opinion and Research* 5, 2 (Spring 2005): 57-64. The World Bank (2008).
- Reinikka, Ritva & Svensson, Jakob, 2007. The Returns from Reducing Corruption: Evidence from Education in Uganda. CEPR Discussion Papers 6363, C.E.P.R. Discussion Papers.
- 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.
- 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.

Appendices

Appendices should be numbered according to the chapter they relate to (i.e. Appendix 1.1 relates to Chapter 1, Appendices 2.1 and 2.2. relate to Chapter 2, etc.).

The numbering is only indicative but the appendices should normally include:

Appendix 1.1: Authorship of this report

Information will be taken from box on first page, including:

Details of Advisory Group membership

Details of Review Group membership

Acknowledgements

Conflicts of interest

Appendix 2.1: Inclusion and exclusion criteria

(1) *The evaluations have to have taken place in a country classified as a “lower, lower middle, or upper middle income” by the World Bank at the time the intervention being studied was implemented.*

(2) *The evaluations have to assess the impact of eliminating primary or secondary public or private school fees [including, tuition fees, fees for textbooks and other instructional materials, compulsory uniforms, Parent Teacher Association (PTA) dues, community contributions to district education boards, and other school-based activity fees (exam fees, etc.)] whether at the universal level or the community/group, district, school, or grade/age level. We will collect all outcome measures reported in the evaluations, such as impact on enrolment, persistence, achievement, and school quality, and any cost-benefit information provided.*

(3) *The evaluation study report is published or available through December 2009, without regard to language or publication type.* In concert with Campbell principles, we will attempt to find English and non-English studies. In addition, we will include published and unpublished studies (e.g., from conference papers, dissertations, technical reports).

Our search will focus on identifying:

(4) *Randomized controlled trials or quasi-experimental evaluations with some evidence that the groups being compared are equivalent.* Our review includes evaluations that randomly assign entities (at any level) to intervention or control conditions. However, because randomization is not possible in certain evaluation scenarios (e.g., in retrospective evaluations or when the program has already been implemented), we will also include evaluative studies that use quasi-experimental designs, provided that they offer evidence that the comparison groups were equated. Such quasi-experiments include regression discontinuity design, propensity scores, covariate matching, and other pre or post intervention matching. We include both types of studies and will include study design as a moderator in our later analyses (see discussion on moderator variable analysis in later section).

In addition, we will include:

(5) *Non-experimental and quasi-experiments without pretest group equivalency* will also be captured and examined, but not included in effect size estimates.

(6) *High-quality descriptive quantitative studies.* We will identify and examine non-causal descriptive quantitative studies, as long as they include sufficient methodological detail as to be replicable, but will not include effect sizes for these studies.

(7) *Qualitative studies.* We will also examine qualitative studies that shed light on possible theory, implementation and context issues, but will not include them in effect size estimates.

Appendix 2.2: Search strategy for electronic databases

Keyword strategies for bibliographic databases

The databases in Appendix 2.3 can be somewhat idiosyncratic. Thus, we believe the best strategy is to conduct a broad search of the available databases that errs on the side of sensitivity rather than specificity. In other words, we would rather get many titles and abstracts to sift through rather than potentially miss relevant citations because our search terms were drawn too narrowly. We will use two search strategies to identify relevant studies. The first strategy will identify studies conducted in low-income developing countries using the following search terms:

Afghanistan* OR Samoa* OR OR OR OR Bangladesh* OR Pakistan* OR Bengal* OR Benin* OR Dahomey* OR Burkin* OR Volta OR Volta* OR Burundi* OR Cambodia* OR Khmer* OR Kampuchea* OR Africa* OR Chad* OR Comoros* OR Congo* OR Zaire* OR Somali* * OR Eritrea* OR Ethiopia* OR Gambia* OR Ghana* OR "Gold Coast" OR Guinea* OR Haiti* OR Kenya* OR Korea* OR Kyrgyz* OR Lao* OR Liberia* OR Madagascar* OR Malaw* OR Nyasaland* OR Mali* OR Mauritania* OR Mozambique* OR Myanmar* OR Burm* OR Niger* OR Rwanda* OR "Sierra Leone" OR "Solomon Islands" OR Tajikistan* OR Tanzania* OR Togo* OR Uganda* OR Zambia* OR Zimbabwe* OR Asia* 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 "impoverished country" OR "impoverished region".

Then, to identify studies that evaluate interventions involving the elimination of school fees, we will add the following search terms: fee* OR levies OR payment* OR tuition OR abolish OR eliminate OR exemption OR “free basic education” OR FBE OR “School Fee Abolition Initiative” OR SFAI OR “free primary education” OR FPE OR “free basic education” OR

FBE OR “universal primary education” OR UPE “universal free primary education” OR textbook* OR due* OR contribution*

Such searching is an iterative process and we will modify as we retrieve studies. This strategy may produce a number of false positives, but our experience is that examining the abstracts is not time consuming and researchers can go through them quite quickly.

Wherever possible, we will limit our searches by a descriptor that indicates the grades that our review is targeting: K-12 (for example, selecting primary and secondary education, or elementary, middle, and high schools). This would have the advantage of screening out preschool and college age studies.

If the database is not focused on education (e.g., *Sociological Abstracts*), the above strategy must then be supplemented by something that identifies educational research. In some databases, that will be a classification code; for example, in *Sociological Abstracts (Sociofile)*, one can limit the abstracts to those dealing with “sociology of education”. But in many of others, there is no classification code. Whether classification codes exist or not, we will use truncated versions of keywords related to the educational outcomes of interest, such as dropout, attendance, and enrollment/enrolment to try to reduce the number of false positives. If there are geographic descriptors for country or region, we will incorporate that into our search process.

We recognize that specific search strategies may have to be developed for each database. What works in identifying potential studies in ERIC will not work in searching World Bank Documents. The appendices to our final review report will carefully document all keywords used for each database to permit replication.

Appendix 2.3: List of bibliographic databases to be searched

- Australian Education Index
- British Education Index
- British Library
- CBCA Education
- Campbell Collaboration Social, Psychological, Educational and Criminological
- Campbell Collaboration Prospective Trials Register (C2-PROT)
- EBSCO Megafile
- Econlit
- Education Index
- Eldis
- ERIC
- 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
- JOLIS (World Bank, International Monetary Fund, International Finance Corporation)
- Medline
- PAIS Archive

- PAIS International
- Periodical Contents Index
- PolicyFile
- Proquest Dissertations and Theses
- Sage Family Studies Abstracts
- Sage Journals Political Science Full-Text Collection
- Sage Journals Sociology Full-Text Collection
- Sage Urban Studies Abstracts
- Social Service Abstracts
- Social Service Research Network
- Social Work Abstracts
- Sociological Abstracts
- UNESCO (UNESDO and UNESBIB)
- Worldwide Political Science Abstracts

Appendix 2.4: 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*.

**DFID Review: Impact of the Abolition of School Fees in
Low-Income Developing Countries
CODING INSTRUMENT**

Coder:

- Claire Morgan
- Anthony Petrosino
- Trevor Fronius
- Other _____

I. RESEARCHER AND STUDY CHARACTERISTICS

What year was the primary document published? _____

What was the type of document?

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

What was the type of study?

- RCT
- QED with equating of groups
- QED without equating of groups
- Non-experiment/Descriptive quantitative
- Descriptive qualitative
- Other

What was the methodology used? _____

What was the type of intervention?

- Universal free primary education
- Elimination of all school fees for targeted group
- Elimination of tuition fees for targeted group
- Elimination of uniform fees for targeted group
- Elimination of textbook fees for targeted group
- Other _____

Who implemented the policy/intervention?

- Government
- NGO
- Other _____

What is the quality of the study methodology?

- High
- Medium
- Low

In what country did the evaluation take place? _____

What was the setting for the evaluation? _____

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

What other information was provided on the context for the evaluation?

Baseline enrolment data: Males _____ Females _____

(Use enrollment 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
- 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
- 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/Neighborhood
- 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 Enrollment (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 enrollment (None/Low/Moderate/High)

III. INTERVENTION AND CONTROL CONDITIONS

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 _____

Age/school level/grade _____

Percentage of participants that were female _____

Poverty/SES (indicate currency Pls using if providing income/wages)

V. OUTCOMES

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

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

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 (include notes on scale and sustainability of intervention)