

SAGE Research Methods Cases

Dropout Prevention Studies in Developing Countries: A Systematic Review and Meta-Analysis

Contributors: Anthony Petrosino & Claire Morgan & Trevor Fronius & Emily E. Tanner-Smith & Robert F. Boruch

Book Title: SAGE Research Methods Cases

Chapter Title: "Dropout Prevention Studies in Developing Countries: A Systematic Review and Meta-Analysis"

Pub. Date: 20140512

Access Date: March 03, 2015

Publishing Company: SAGE Publications, Ltd.

City: London

Print ISBN:

Online ISBN: 9781446273050

DOI: <http://dx.doi.org/10.4135/978144627305013510256>

©2014 SAGE Publications, Inc. All Rights Reserved.

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/978144627305013510256>

Dropout Prevention Studies in Developing Countries: A Systematic Review and Meta-Analysis

Link to the Research Output

- <http://www.campbellcollaboration.org/lib/download/2393/>

This case illustrates ideas about systematic reviews of evidence and meta-analysis in a stepwise fashion. Presented is a discussion about studies of the effectiveness of interventions that are designed to prevent school dropout in low and low to middle income countries and to use tests in high-quality randomized controlled trials and high-quality quasi-experiments. The coverage includes framing the questions, searching literature, computing effect sizes, and portraying results on 'what works'.

Learning Outcomes

Readers will be able to

- Describe a high-quality systematic review and associated meta-analysis and their ingredients
- Interpret a forest plot of effects sizes in a meta-analysis
- Become more informed consumers of systematic reviews and meta-analyses in the future

Meta-Analysis: Introduction

In Saint Paul's epistle to the Thessalonians, he declared that one ought to try all things and hold firmly to that which is true. The thoughtful reader can find the same idea embodied in the Koran, in the writings of the Buddha and Taoist thinking, and elsewhere.

Scientists with a taste for understanding how to synthesize studies of the impact of social programs are kindred in spirit though they may not refer to the sages. Nowadays, with a multitude of evidence of highly varying quality and bold claims about the purported effectiveness of programs, finding out what is of value is arguably not simple. Meta-analysis is a way of arranging how one might do so. The aim is to summarize, in statistical terms, what efforts are effective and which ones are not, based on systematic reviews of what has been tried out and judging the quality of relevant evidence based on high standards.

This case illustrates meta-analysis in a stepwise fashion and in the context of interventions that tested to understand what works to prevent school dropout in developing countries. The steps begin with framing the questions to be addressed and proceed in an orderly way, through literature searches, study screen, computation of effect sizes, and producing results.

Objectives in This Meta-Analysis Case: Developing Countries

Framing the questions to be addressed in any systematic review and meta-analysis is a crucial first step. For this project, our objectives were to respond to the following primary and secondary questions on the context of low and middle income countries:

- *Primary question:* What are the effects of interventions implemented in developing countries on measures of students' enrollment, attendance, graduation, and progression?

- **Secondary question:** What are the effects of interventions on learning outcomes as measured by students' test scores, grades, and other achievement measures?

The rationales for the questions are important. They lie in contemporary policy concerns about education issues in developing countries, elaborated in by the World Bank (2004, 2008), Birdsall, Levine, and Ibrahim (2005), Duflo and Kremer (2005), and others. School dropout rates by the age of 12 approach 70% in many such countries.

The approach taken here employs the best of contemporary methods for systematic acquisition and review of high-quality studies of the topic and the statistical meta-analysis of the results of these studies. Each of the steps in the process, and the application of methods, is summarized below. More detailed information on particular steps and technical nuances can be found in the report produced by Petrosino et al. for the international Campbell Collaboration's web site at <http://campbellcollaboration.org>.

Study Selection

Criteria for Including and Excluding Studies in This Review

Excellent meta-analyses include explicit criteria for inclusion and exclusion of studies that are considered. Studies were included here if they met the eligibility criteria below, including at least one quantifiable outcome relevant to the primary question. If they satisfied that criteria, we then examined these same studies for quantifiable measures relevant to the secondary question on learning outcomes.

For this project, only evaluation studies that had the following characteristics were included. They

- assessed the impact of an intervention that included primary or secondary school outcomes (Kindergarten-12th grade in the US context) relevant to the primary research question;
- used a randomized controlled trial (RCT), or a quasi-experimental design (QED) approach in which a strong case for baseline control was made;
- were conducted in a country classified as a 'low or middle income nation' by the World Bank at the time the intervention under study was implemented;
- included at least one quantifiable outcome measure of school enrollment;
- were published or made available before December 2009, without regard to language or publication type;
- included data on participants from 1990 or beyond.

For instance, the Progresa/Opportunities experiment in Mexico exemplifies a study that was included in our review. In this study, cash transfers conditional on child attendance in school (and other conditions) were provided to households in experimental areas that were randomly allocated to the cash transfer intervention or to a control condition. Numerous reports have indicated that Progresa had positive impacts on school enrollment and attendance (e.g. Schultz, 2004). Other studies, however, demonstrated negligible or more modest effects or even negative effects. They too have been included in our review of course.

The second criterion is especially important in view of the problems engendered by poorly designed studies of the impact of related interventions. See, for instance, Boruch (1997), Newman et al. (1995), and Bloom, Michalopoulos, Hill, and Lei (2002) among others.

Search Strategy for Identifying Relevant Studies

Making clear how one searched the literature, and exploiting and identifying all the resources available, is a hallmark of high-quality meta-analyses. Our goal in this project was to identify relevant reports in both published and unpublished literature. Many of the databases that were used include the grey or fugitive literature, that is, documents

that are not published in journals or books but are found in hard-to-get sources like dissertations, theses, government and technical reports, obscure trade journals, and conference papers. We included searches of the World Wide Web, because such searches can identify reports that are made available at web sites but are not published in journals. Direct contacts with colleagues were designed to get at more of the grey literature. To accomplish our goal, we used five major strategies:

- *Electronic searches of bibliographic databases.* Researchers searched available online resources and databases at the University of Pennsylvania and Bridgewater State College, including Education Resources Information Center (*ERIC*), *PAIS International/Archive*, *Sociological Abstracts*, and *World Bank Documents*. A complete list of databases that were searched and a full listing of the keywords used in each database are given at <http://campbellcollaboration.org>.
- *Hand searches of relevant journals.* Because electronic searches can miss relevant studies, we ‘hand searched’ (i.e. visually inspected the table of contents and the articles) five journals: *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 searched by hand every issue of each of the five journals through 2009.
- *Citation chasing.* The reference section of every retrieved report was checked to determine whether any possible eligible evaluations were listed. We checked references for reports that were retrieved that were ineligible for this review, including syntheses, nonexperimental studies, methodological papers, and descriptions of policies or interventions.
- *Contacting the ‘informal college’ of researchers in this area.* There is a network of researchers involved in conducting experimental and quasi-experimental studies relevant to developing nations. We identified the lead authors of such studies or relevant documents (e.g. reviews, nonevaluative studies), identified their e-mail addresses, from a Google search of the World Wide Web, and e-mailed them query letters.
- *Internet Searches and specialized holdings.* We also used the ‘advanced search’ options in *Google* and *Google Scholar* for broad searches of the World Wide Web. This was supplemented by specialized searches of specific web sites that could reference relevant holdings such as the Center for

Population Development and Activities, the Massachusetts Institute of Technology's Poverty Action Lab, Yale University's Innovations for Poverty Center, the National Bureau for Economic Research (NBER), the Network for Policy Research, Review and Advice on Education and Training (<http://www.norrag.org>), the Network of Networks Impact Evaluation Initiative (NONIE), the World Health Organization (e.g. their Annotated Bibliography of Selected Research on Civil Society and Health), the Organisation for Economic Co-operation and Development (OECD), and 3ie's own database of impact evaluations.

Keyword Strategies for Bibliographic Databases

High-quality meta-analyses and systematic reviews are attentive to keywords used to search bibliographic databases and report the strategies utilized. Our search strategies here were of two major types. First, and for most databases, we developed a long list of keywords to identify three major study eligibility criteria: (1) keywords relevant to developing nations; (2) keywords relevant to the outcomes of enrollment, dropout, persistence, and so on; and (3) keywords relevant to experimental and quasi-experimental evaluations. These were used successfully in most databases; in a few instances, however, the yield was still so large that we instituted a (4) fourth criterion of keywords relevant to youth. Such searching is an iterative process. We modified our terms as we retrieved studies. The second search strategy focused on databases that did not permit complex searches. In these, we searched by using one or a few keywords at a time.

Retrieving and Final Screening of Studies

Full text reports must be acquired to determine study eligibility for review. When a full text report was received, we read it to ensure that it met the minimum evaluation design criteria above and included at least one outcome of school enrollment. A decision was

made by one or more of the first three authors to include the study tentatively in the review.

Study Coding

Extracting and Coding Information from Each Study

Putting diverse kinds of relevant information into well-defined categories is a part of science and an essential step in classifying the kind of evidence that is at hand in a meta-analysis. We designed a coding instrument to guide us in extracting information from each study (see Appendix 8.5 at <http://www.campbellcollaboration.org/lib/download/2393/>). The instrument contains items that describe the researcher's field or discipline, publication type and year published, context (country and classification of economy), evaluation design (whether RCT or QED), methodological quality (how the study handled statistical bias, etc.), treatment and control or comparison group, participants (e.g. grade), and outcome data (i.e. on enrollment and learning outcomes).

Coding Reliability

No one is perfect when it comes to reviewing evidence. Hence, multiple reviewers must be used and the agreement among them must be established. To ensure that we achieved good inter-rater reliability, the first three coauthors in this study read and recorded information from a random sample of reports (12% or 17% of the final sample). Items with lower rates of agreement (less than 80%) were investigated to determine the source for conflict.

Criteria for Determination of Independent Findings

In the courts of many countries, and as in ordinary life, and as in statistical meta-analyses, observations from independent sources are crucial in understanding whether the evidence at hand is dependable. Our criteria for handling studies were as follows:

- We identified only one effect size per analysis, bearing on the main outcome variable, so as to avoid the problems engendered by data mining and multiplicity of statistical tests.
- Our unit of analysis was the evaluation study, not the evaluation report.
- An evaluation study was considered distinct if it used a different sample.
- The primary analysis or study design was the ‘most rigorous’, or one that provided the most controls.
- Overall versus subgroup effects were the focus of attention.
- Individual level effects rather than aggregate/institutional were the focus where possible.

Statistical Procedures and Conventions

Sophisticated meta-analyses of the results of multiple studies depend on standardized indicators of the effectiveness of the interventions that are field-tested, estimates of the variance of the effect sizes, and computation of average effect sizes all summarized in forest plots. Here, standardized mean differences (Cohen's d) were used as the effect size metric for all the primary and secondary outcomes of interest. These are appropriate for measuring group differences in mean levels of a continuously measured outcome (Lipsey & Wilson, 2001; Wilson, 2011). All effect sizes were coded so that positive effect sizes represented better outcomes (e.g. higher enrollment, lower dropout). Standardized mean difference effect sizes were calculated as

$$d = \frac{\bar{X}_{TX} - \bar{X}_{CT}}{s_{pooled}}$$

where the numerator is the difference in group means for the intervention and control groups, and the denominator is the pooled standard deviation for those groups. The variance of the standardized mean difference effect size was calculated as

$$V_d = \frac{n_{TX} + n_{CT}}{n_{TX} \times n_{CT}} + \frac{d^2}{2(n_{TX} + n_{CT})}$$

The n values represent the sample sizes used in intervention (subscripted TX) and control groups (subscripted CT). We used Comprehensive Meta-Analysis (CMA) algorithms to estimate effect sizes and variances and to statistically combine results from the evaluations.

Forest plots generated by CMA were used in what follows to display the results from the effect sizes. A forest plot illustrates the effect size and confidence intervals (precision) for each study and also provides the combined effect across all studies. To interpret the forest plot, the little diamonds represent the effect size and the band around them represents the confidence interval (the band represents the range of effects that the effect size falls within, usually within 95% probability; note that smaller is better because it means the effect size is more precise). The line down the zero of the forest plot represents a ‘zero’ effect, that is, both groups had the same exact outcome. Effect sizes to the left are negative, meaning that the intervention had a harmful outcome (the treatment group did worse than the control group), and effect sizes to the right are positive, meaning that the treatment group performed better than the control group. All meta-analysis software such as CMA and many statistical programs such as Stata will easily produce forest plots to graphically portray meta-analysis outcomes.

We report overall effects across all interventions on the four major outcomes (enrollment, attendance, dropout, and progression) and on the four types of learning outcomes (math, language, global tests, and other achievement measures). [Figure 1](#) presents how we organized outcomes across the studies.

Figure 1. Organization of outcomes.

Primary Outcomes (mandatory)			
Enrolment Participation	Attendance	Dropout	Progression
Enrolment Participation	Attendance Absenteeism	Dropout Retention Completion Graduation	Progression to Next Grade Repetition

Secondary (if study had at least one primary outcome only)			
Math Scores	Language Scores	Global Test Scores	Other Achievement
Math	Reading Verbal Native Language English	National Test Scores Average Test Scores Oral/Written Scores	Grades Classes: Passed/Failed

We then descriptively examine a number of moderators. These moderators are approached and interpreted descriptively rather than statistically, as they are often based on small numbers of studies (the ‘small cell’ problem), and such analyses can be significant by chance if large numbers of variables are considered (the ‘capitalizing on chance’ problem). Our analyses examined

- Intervention types,
- RCTs versus QEDs,
- World Bank classification of economies,
- School level,
- Gender.

Pipeline of Studies

Not all studies initially deemed eligible review provided enough information for meta-analysis. A total of 116 RCTs and quasi-experiments met our initial screening for eligibility. All of these were coded for study characteristics but not for effect sizes. We then conducted a second screening to again ensure that there was evidence of baseline control and that the original studies reported on participant data from 1990 or after (eligibility criteria 6). This left 81 studies. Eight studies, as mentioned earlier, did not report outcome data that we could use to create a quantifiable effect size. This left 73 total studies in our final sample.

Descriptive Statistics

Characterizing the final sample of studies involves simple statistics. In this case, the sample of studies is diverse. They were conducted in 27 different nations, with Kenya (N = 12), India (N = 9), Bangladesh (N = 6), Colombia (N = 5), and Jamaica (N = 5) the most common. Not surprisingly, as [Table 1](#) shows, most studies were conducted in the poorest developing nations (51% in Lower Income Countries or LICs); 34% were conducted in Lower Middle Income Countries (LMICs). All of the nations identified above except Colombia (which is defined as an Upper Middle Income Country (UMIC) economy) fall into those two classifications.

Table 1. Comparison of average effects for studies reporting individual level data and those reporting only aggregate level data.

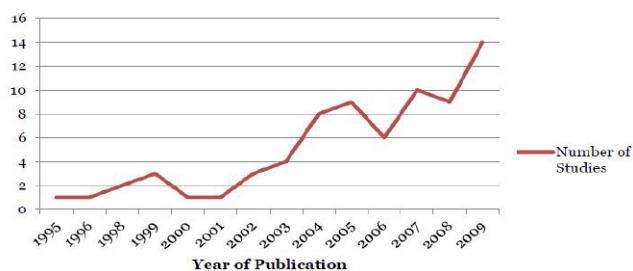
	Standardized mean effect (d)	Variance (v)
Individual N's	0.161	0.0026
Aggregate N's	0.199	0.0067

Nearly 40 different kinds of interventions were tested across these 73 studies; broadly, Conditional Cash Transfers (N = 13), funding or grants to communities (N = 5), school breakfasts or lunches (N = 5), or remedial education or tutoring (N = 5) were the most common. Most of these programs targeted primary school-aged children (N = 44, 60%), with 10 focusing exclusively on secondary school-aged children (14%). A larger minority of studies involved interventions that included both primary and secondary students (N = 19, 26%). Nearly 9 in 10 studies (N = 65, 89%) included both boys and girls in the intervention; the remainder focused exclusively on girls (N = 8, 11%).

In all, 52 of the studies (71%) used randomization to assign participants to groups, and 21 (29%) used quasi-experimental procedures. This reflects the strict processes we used to select quasi-experiments, and the recent 'randomized revolution'. These studies

were published from 1995 to 2009. As [Figure 2](#) indicates, the increase in the number of eligible studies since the early 2000s has been remarkable.

Figure 2. Number of included studies by year of publication.



These studies assigned individuals and larger aggregate units to treatment and control conditions. Most common were studies that assigned schools to treatment or control conditions ($N = 31$; 43%), followed by the assignment of individuals to treatment and control conditions ($N = 14$, 19%). Some studies assigned villages to treatment and control conditions ($N = 13$, 18%).

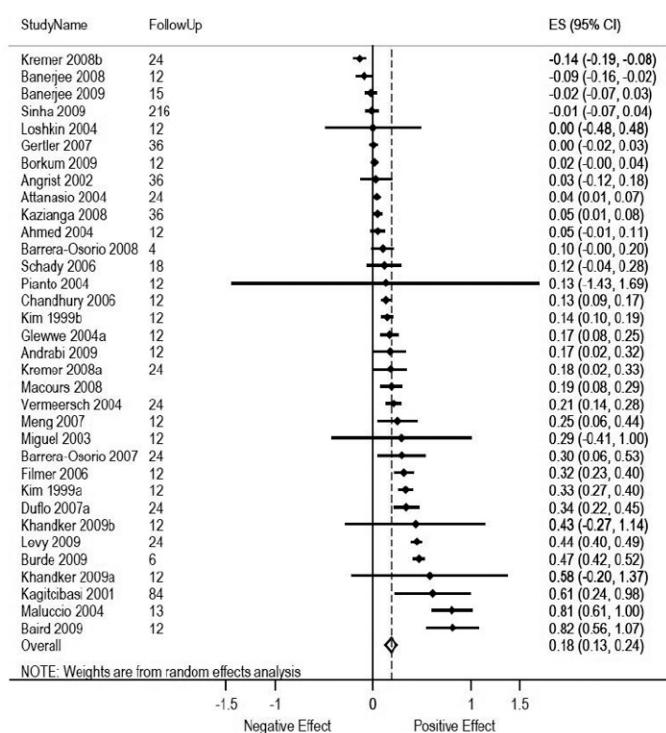
Meta-Analyses

Statistical meta-analysis and graphical portrayal of results are straightforward, given recent advances in the field and readily available software such as CMA. Using inverse variance random effects weights, we estimated the overall mean effect size d across studies, separately for the different types of outcomes. Cohen's d (also referred to as a standardized mean effect) is interpreted in the analyses as positive if there is a positive impact for intervention (e.g. if enrollment increased or dropout decreased), negative, such as $-.10$, if there was a negative impact (e.g. an increase in dropout or decrease in enrollment), and 0 if the effect for the intervention was identical for the treatment group and the control group (e.g. 95% enrollment rate in both groups). Generally speaking, an effect size estimate of $.10$ reflects 1/10 standard deviation improvement for treatment participants compared to control participants. Here, we employ forest plots to portray some of the results of each study and of the meta-analyses.

Overall Intervention Effects, Primary Analyses of Enrollment, and Other Outcomes

Here are the average effects of the interventions on primary enrollment and other relevant outcomes (attendance, dropout, and progression). In [Figure 3](#), the results are presented for 34 studies that measured the impact of an intervention on an enrollment outcome.

Figure 3. Main effects on school enrollment (n = 34).

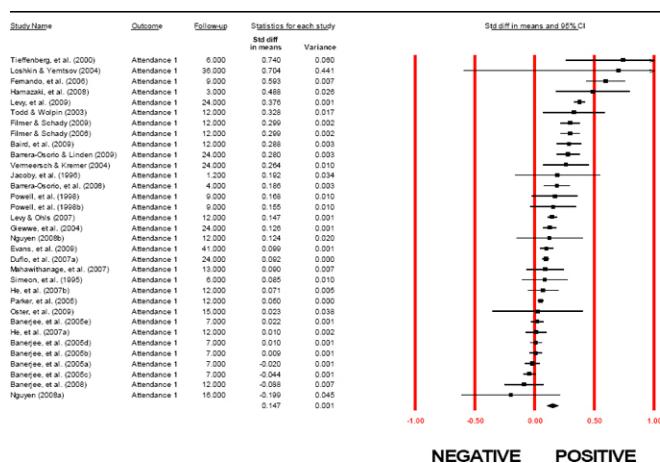


This analysis includes only the first effect reported in the study; as indicated, the first follow-up measure ranges from 4 to 216 months. Collectively, the average treatment

effect is positive ($d = .18$) and ranges from -.14 to .82. Only five studies reported a negative impact on enrollment (to the left of zero).

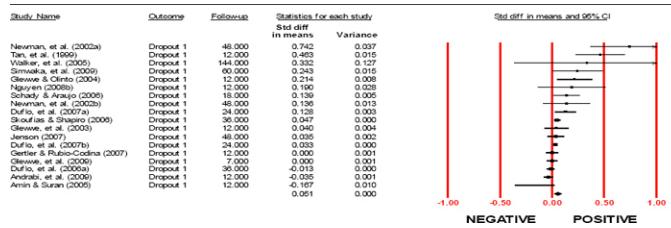
A total of 33 studies included at least one quantifiable outcome of attendance in their analyses. These studies measured attendance at a wide range of time intervals, ranging from 1.2 to 41 months. [Figure 4](#) presents the effect sizes for these studies. Similar to enrollment results, the overall effect is positive ($d = .15$), ranging from -.20 to .74. Four studies reported negative results on attendance (to the left of zero).

Figure 4. Average intervention effect on attendance (N = 33).



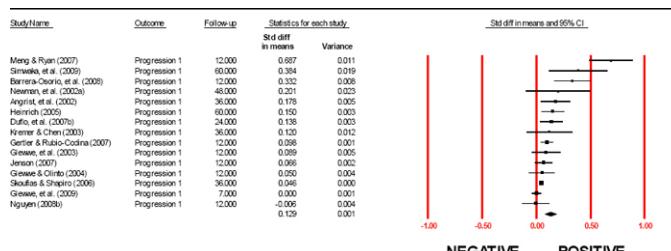
There were 18 studies that included at least one quantifiable outcome of school dropout in their analyses. Follow-up of outcomes for dropout varied greatly across these 18 studies, ranging from 7 to 144 months. [Figure 5](#) presents the effect sizes for these studies. Compared to enrollment and attendance, the overall effect is positive but smaller ($d = .05$), ranging from -.17 to .74. Three studies reported negative effects, that is, an increase in school dropout (to the left of zero).

Figure 5. Average intervention effect on school dropout (N = 18).



There were 15 studies that included at least one quantifiable outcome of progression. Follow-up of outcomes for progression varied greatly across these studies, ranging from 7 to 60 months. Figure 6 presents the relevant effect sizes. The overall effect is positive and similar to those reported for enrollment and attendance ($d = .13$), ranging from -.01 to .69. Only one study reported negative effects (to the left of zero) on progression in school.

Figure 6. Average intervention effect on progression in school ($N = 15$).



Moderating Variable Analyses

It is likely that the effect sizes vary because some meta-analyses involve a large number of studies and diverse interventions, samples, and settings. Examining how effect size varies with moderating variables is good practice. In this case, we examined five moderating variables: type of intervention, World Bank economy classification, type of evaluation design, whether the intervention targeted females, and whether the intervention included primary or secondary school students, or both.

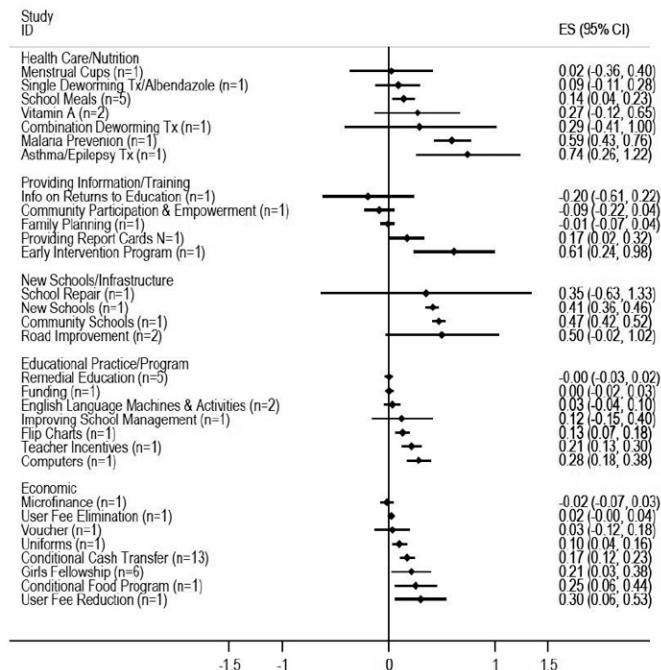
For these analyses, all 73 studies were included once. Because not all studies measured the same outcomes, the effect size in these moderating variable analyses

is based on the first effect reported in the study on one of the four primary outcomes (enrollment, attendance, dropout, and progression). If a study reported more than one effect on one of these four outcomes, we selected one based on the following order: enrollment, attendance, dropout, and, finally, progression. This order does not reflect a hierarchy of importance, rather the fact that most studies reported effects on enrollment or attendance.

Specific Intervention Type

One important policy and practice question is whether effect sizes vary across specific intervention types. [Figure 7](#) presents the results across 36 interventions. The interventions are listed in alphabetical order, and the number of studies in each intervention category is in parentheses. Most interventions are positive in direction. The five interventions reporting the largest average effects on a primary outcome are the effects of asthma/epilepsy treatment on attendance ($d = .74$), early intervention on enrollment ($d = .61$), road improvement on enrollment ($d = .50$), a dropout prevention program on school dropout ($d = .46$), and building new schools on enrollment ($d = .45$). Only 6 of the 36 specific interventions reported average negative effects (to the left of zero). These were community participation and empowerment ($d = -.09$), family planning ($d = -.01$), livelihood skills training ($d = -.17$), microfinance ($d = -.02$), remedial education ($d = -.004$), and teacher training ($d = -.01$).

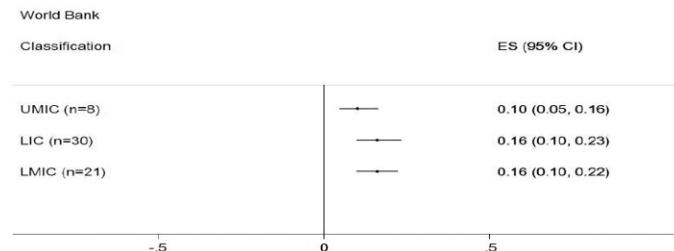
Figure 7. Average effects across specific intervention types.



World Bank Classification of Economies

As mentioned earlier, studies were conducted of interventions implemented across a wide range of developing nations. In this analysis, we examined the average effect for interventions implemented in the three types of developing nations, as defined by the World Bank classification of economies (LIC, LMIC, and UMIC). As [Figure 8](#) indicates, the average effect across the 25 LMICs was the largest ($d = .18$), followed by the average effect across the 37 LICs ($d = .14$). The lowest average effect was for the 11 studies implemented in UMICs ($d = .09$).

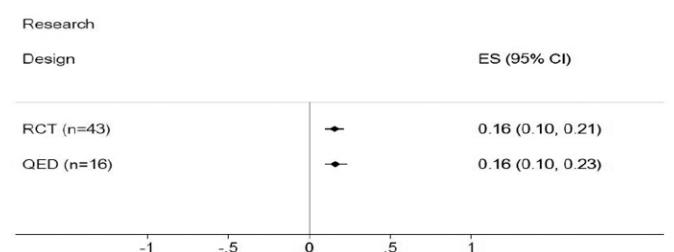
Figure 8. Average effects across World Bank classification of economies.



Type of Evaluation Design

As discussed above, our sample comprises a large majority of RCTs, likely reflecting our stringent eligibility criteria and screening of QEDs. In this analysis, we compare the average effect size for RCTs versus QEDs. As [Figure 9](#) indicates, there was very little difference in these average effects ($d = .15$ for RCTs, $d = .14$ for QEDs).

Figure 9. Average effects for different evaluation designs.

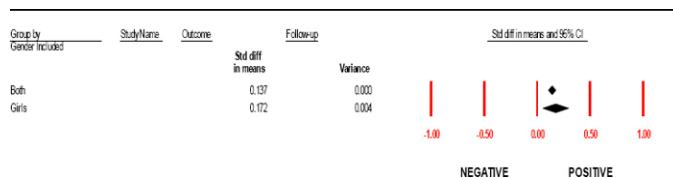


Did Intervention Specifically Target Females?

In recent years, there has been a strong emphasis by donor agencies and the governments of developing nations on specifically targeting females for educational initiatives. In this review, eight studies tested interventions that specifically targeted females (although some may have examined spillover effects on boys), including six that were scholarship/fellowship programs. [Figure 10](#) examines the average effect for

those eight studies and compares it to the average effect for the remaining 65 studies that tested interventions that included boys and girls. As the figure shows, the average effect for female-focused interventions is slightly larger ($d = .17$ to $d = .14$).

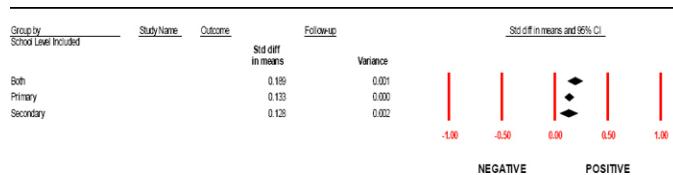
Figure 10. Average effects for studies of interventions targeting females or including both genders.



Did the Intervention Include Primary or Secondary School Students?

Another aspect of the wide diversity of these studies is that some target primary school students, some secondary students, and others include outcomes for students at both school levels. [Figure 11](#) examines the average effects for interventions across these school levels. As the figure indicates, the effect sizes are very similar for studies including only primary student outcomes and those including only secondary student outcomes ($d = .13$ vs $d = .13$). The largest average effects are for those interventions including both types of students ($d = .19$).

Figure 11. Average effects by school level targeted.



Methodological Quality Checks

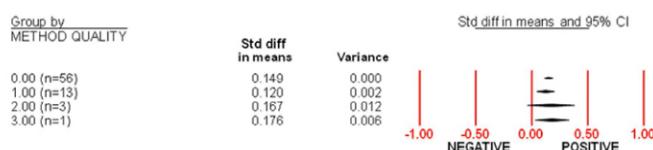
Effect sizes may vary with subtle problems in the study's execution. For each study, we captured information about any issues with crossovers (persons receiving condition they were not assigned to), selection bias (e.g. breakdowns in randomization or unusual unequal distributions in groups), loss of participants due to attrition or database matching issues, and intervention fidelity and implementation issues. The first three authors also rated each study according to their own perceptions of whether the problems presented a threat to the findings reported in the study. These ratings were categorized as 'low', 'moderate', or 'high'. If there were no indicated problems, the threat to the study was rated as 'none'. It should be noted, however, that these ratings are subjective, and they are based entirely on what is reported in the study documents. However, these ratings had good reliability across the three first authors in our check of inter-rater reliability (generally, about 75% agreement across the 12 studies). This indicates that the three authors were identifying the problems and rating the degree of threat to the conclusions in similar fashion.

We examined whether a rating of 'moderate' or 'strong' threat to the study's conclusions on the four methodological items influenced the average effect size across the studies. For example, a study that had none of the four items rated as a 'moderate' or 'strong' threat to validity received a '0'. Likewise, a study that received a rating of a 'moderate' or 'strong' threat on all four of the methodological items was scored a '4'. For this analysis, as with the moderating variable analyses, all 73 studies were included once. The effect size in this analysis is again based on the first effect reported in the study on one of the four primary outcomes (enrollment, attendance, dropout, and progression). If a study reported more than one effect on one of these four outcomes, we selected one based on the following order: enrollment, attendance, dropout, and, finally, progression.

[Figure 12](#) presents the results and is instructive in several respects. First, the methodological problems in the largest majority of studies in this review were rated as presenting little or no threat to study conclusions ($N = 56, 77\%$). Second, only four studies (5%) had two or three methodological problems rated as 'moderate' or 'strong' threats to study conclusions, scoring 2.00 or 3.00 on the Method Quality score. These findings are likely due to the especially strong sample of RCTs and QEDs screened into

this review. However, the results indicate that the small number of studies that scored a 2.00 or 3.00 did report an average effect size larger ($d = .17$ for 2.00, $d = .18$ for 3.00) than studies that reported no threats or one threat ($d = .15$ for 0.00; $d = .12$ for 1.00).

Figure 12. Average effects by rating of study methods quality.



Using Individual versus Aggregate Sample Sizes to Compute Effect Sizes

As mentioned previously, when possible, we used sample sizes for the individual students in the studies rather than the sample sizes for the aggregate units that were randomly or quasi-experimentally assigned to conditions. So, for example, if a study randomly assigned 10 villages each to treatment and control conditions, and then reported analyses on enrollment using individual sample sizes, we used those individual sample sizes to compute effect sizes. All such studies took clustering into account when analyzing at the individual level, and so no corrections for lack of clustering were applied to these data.

We conducted one post hoc methodological check to see how different the average effect sizes were when using aggregate units of assignment versus using the sample sizes of individual students in the studies. [Table 1](#) presents the results of a comparison for 12 studies (16% of the total review sample) in which the effect sizes for aggregate and individual sample sizes were computed. As [Table 1](#) indicates, the differences in the average effects and variances, at least in this analysis, are not substantial ($d = .16$ when using individual sample sizes, and $d = .20$ when using aggregate sample sizes).

Study Conclusion

In this review, we identified 73 experimental and quasi-experimental studies that examined the impact of an intervention on at least one primary outcome of enrollment, attendance, dropout, and progression in a developing nation. We also examined the effects of these interventions on secondary learning outcomes of math scores, language scores, global test scores, and other achievement. [Table 2](#) summarizes the results for these analyses. As [Table 2](#) indicates, average effects across the four primary and four secondary outcomes were positive in direction. Effects on enrollment, attendance, and progression were larger than those reported across studies that measured dropout. For secondary learning outcomes, average effects on math and language scores were larger than those reported across studies that measured global test scores or 'other achievement'.

Table 2. Summary of meta-analyses.

Outcome	Number of studies in analysis	Standardized mean effect (d)
Primary		
Enrolment	34	.184
Attendance	33	.147
Dropout	18	.051
Progression	15	.129
Secondary		
Math	25	.164
Language	25	.180
Global test scores	10	.061
Other achievement	5	.048

These average effects should be tempered by noting the diversity of studies, samples, countries, interventions, and measures in the 73 studies synthesized here. [Figure 7](#)

presented the average effects for the 36 types of interventions evaluated by this sample of studies. They present some early indications, comparatively, about the effects of these interventions on the outcomes summarized here. Several cautions are in order. First, in many intervention categories, only one study has been reported. Second, our analyses focused on primary outcomes of enrollment, attendance, dropout, and progression (and then examined secondary outcomes of learning). There may be other important outcomes for child employment, health, and other school outcomes (e.g. teacher attendance and efficacy) that we have not examined here.

Apart from the average effects for specific interventions captured in [Figure 7](#), we also examined four other moderating variables. These are summarized in [Table 3](#). The table indicates that studies that targeted LMICs included both primary and secondary school student outcomes and delivered an intervention focused on females that reported larger average effects.

Table 3. Summary of moderating analyses.

Outcome	Standardized mean effect (d)
World Bank classification	
LIC	.129
LMIC	.184
UMIC	.091
School level	
Primary	.132
Secondary	.109
Both	.189
Board design	
RCT	.150
QED	.130
Gender	
Boys and girls	.138
Girls only	.149

What Is Learned by Doing Meta-Analysis?

There are several advantages to doing meta-analysis. First, it provides a method for statistically summarizing studies. It would be very difficult to do this without some statistical method, in the same way that it would be hard to summarize 73 police reports or school records without the use of statistics.

Second, meta-analysis provides a transparent and explicit method for reviewing studies to determine ‘what works’. Readers can draw their own conclusions, but there is no guesswork involved in how we identified studies, how we found them, what we did to create effect sizes and analyze them, and why we draw the conclusions presented. Contrast this with the traditional reviews that were done before meta-analysis—often

published by the substantive experts in an area—that often did not provide any details about the methods used to identify what works. In some sense, meta-analysis has moved us from content experts saying ‘trust me’ to researchers saying ‘here’s what I did, test me’.

A third advantage is that meta-analysis allows us to identify an average effect across intervention studies (e.g. [Table 2](#)), and to also analyze how that effect varies depending on changes in context or program. The moderators in [Table 3](#) are a good example of that. This would be very difficult, if not impossible, without meta-analytic statistics.

Finally, this meta-analysis can now be updated periodically to determine how results from new studies impact the findings presented here.

But all is not rosy. Our meta-analysis identified some challenges. Although all of the studies present ‘fruit’, there are some distinctions between the types of interventions that were included in the analysis. They represent, as many critics have stated in the past, apples and oranges (and maybe a few lemons). Thus, some may object to a lumping of these studies together in a single meta-analysis.

There were also some difficulties in creating effect sizes from studies that used econometric or complex statistical models. We were explicit in how we handled these, but there was some uncertainty at times about the best way to go about this. In some sense, meta-analysis can produce an ‘air of orderliness’ even if the studies are messy.

Moreover, some of the programs studied in this sample were actually comprehensive initiatives focused on many outcomes. Thus, a program may have wonderful effects on health or child labor. But by focusing solely on dropout, enrollment, and other learning outcomes, an intervention may appear to be ineffective because it has little discernible effect on education outcomes. This could convince policymakers not to invest in the intervention—even though its full range of benefits was never examined in this review.

Exercises and Discussion Questions

- In a Google search for information about parent education programs, you uncover a site that identifies Model Programs. It provides a description of each program. What questions would you pose about evidence to support the idea that the programs are indeed ‘models’?
- If asked by your boss to describe a meta-analysis, how would you do so?
- What databases or other sources would you search for evidence-based programs in your interest area? Why? And how would you decide that your search is sufficient?
- Develop a list of keywords for a search of evidence-based programs in your interest area, and apply it in a first stage search. Report back on the results and how you might improve the keyword list to assure that your search can be improved.
- Develop a forest plot based on fictitious data, and explain its ingredients and what the plot means to the viewers of the plot.
- What would you advise the readers of your meta-analysis to do if your meta-analysis showed that no well-tested new program worked better than a conventional program to which it was compared?

Web Resources

The Campbell Collaboration. Available from <http://www.campbellcollaboration.org>

The Cochrane Collaboration. Available from <http://www.cochrane.org>

The United States Agency for International Development. Available from <http://www.usaid.gov>

The World Bank. Available from <http://www.worldbank.org>

References

- Birdsall, N., Levine, R., & Ibrahim, A. (2005). *Toward universal primary education: Investments, incentives, and institutions*. London, England: United Nations Development Programme.
- Bloom, H., Michalopoulos, C., Hill, C., & Lei, Y. (2002). *Can non-experimental comparison group methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs?* New York, NY: Manpower Demonstration Research Corporation, 2002.
- Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide*. Thousand Oaks, CA: SAGE.
- Brembeck, C. S. (1962). *Education for national development*. *Comparative Education Review*, 5, 223–231. <http://dx.doi.org/10.1086/cer.1962.5.issue-3>
- Duflo, E., & Kremer, M. (2005). *Use of randomization in the evaluation of development effectiveness*. In G. K. Pitman, O. N. Feinstein, & G. K. Ingham (Eds.), *Evaluating development effectiveness* (pp. 205–247). New Brunswick, NJ: Transaction.
- Filmer, D. (1999). *The structure of social disparities in education: Gender and wealth (Policy Research Report on Gender and Development, Working Paper Series No. 5)*. Washington, DC: The World Bank.
- Filmer, D. & Schady, N. (2006). *Getting girls into school: Evidence from a scholarship program in Cambodia*. Washington, DC: The World Bank.
- Glazerman, S., Levy, D. M., & Myers, D. (2003). *Nonexperimental versus experimental estimates of earnings impacts*. *The Annals of the American Academy of Political and Social Science*, 589, 63–93. doi: <http://dx.doi.org/10.1177/0002716203254879> <http://dx.doi.org/10.1177/0002716203254879>
- Glewwe, P., & Kremer, M. (2006). *Schools, teachers, and education outcomes in developing countries*. In E. Hanushek, & F. Welsh (Eds.), *Handbook on the*

economics of education (Vol. 2, pp. 945–1018). London, England: North-Holland. <http://dx.doi.org/10.1016/S1574-0692%2806%2902016-2>

Glewwe, P., Kremer, M., & Moulin, S. (2009). *Many children left behind? Textbooks and test scores in Kenya*. *American Economic Journal: Applied Economics*, 1, 112–135. doi: <http://dx.doi.org/10.1257/app.1.1.112> <http://dx.doi.org/10.1257/app.1.1.112>

Hanushek, E. A. (1995). *Interpreting recent research on schooling in developing countries*. *World Bank Research Observer*, 10, 227–246. doi: <http://dx.doi.org/10.1093/wbro/10.2.227> <http://dx.doi.org/10.1093/wbro/10.2.227>

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

Kremer, M., & Alaka, H. (2009). *Improving education in the developing world: What have we learned from randomized evaluations?* *Annual Review of Economics*, 1, 513–542. doi: <http://dx.doi.org/10.1146/annurev.economics.050708.143323> <http://dx.doi.org/10.1146/annurev.economics.050708.143323>

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.

Lockheed, M. E., Vail, S. C., & Fuller, B. (1986). *How textbooks affect achievement in developing countries: Evidence from Thailand*. *Educational Evaluation and Policy Analysis*, 8, 379–392. doi: <http://dx.doi.org/10.3102/01623737008004379>

Newman, J., Laura, R., & Paul, G. (1994). *Using randomized control designs in evaluating social sector programs in developing countries*. *World Bank Research Observer*, 9, 181–201. doi: <http://dx.doi.org/10.1093/wbro/9.2.181> <http://dx.doi.org/10.1093/wbro/9.2.181>

Oliver, S., Bagnall, A. M., Thomas, J., Shepherd, J., Sowden, A., White, I., ... Garrett, Z. (2008). *RCTs for policy interventions? A review of reviews and meta-regression*. Retrieved from http://www.pcph.bham.ac.uk/publichealth/methodology/docs/publications/JH09_Oliver_2008_RCTs_for_policy_interventions.pdf

Petrosino, A., Morgan, C., Fronius, T. A., Tanner-Smith, E. E., & Boruch, R. F. (2012). *Interventions in developing nations for improving primary and secondary school enrollment of children: A systematic review*. *Campbell Systematic Reviews*, 19. doi: <http://dx.doi.org/10.4073/csr.2012.19>

Rosenthal, R., & Rubin, D. B. (1982). *A simple general purpose display of magnitude and experimental effect*. *Journal of Educational Psychology*, 74, 166–169. doi: <http://dx.doi.org/10.1037/0022-0663.74.2.166> <http://dx.doi.org/10.1037/0022-0663.74.2.166>

Schultz, T. P. (2004). *School subsidies for the poor: Evaluating the Mexican PROGRESA poverty program*. *Journal of Development Economics*, 74, 199–250. doi: <http://dx.doi.org/10.1016/j.jdeveco.2003.12.009> <http://dx.doi.org/10.1016/j.jdeveco.2003.12.009>

The World Bank. (2004). *World development report 2004: Making services work for poor people*. Washington, DC: Author.

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

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

Wilson, D. B. (2011). *Calculating effect sizes*. Presentation at the CEBCP-Campbell Collaboration Joint Symposium on Evidence Based Practice. Fairfax, VA: The Campbell Collaboration. Retrieved from http://www.campbellcollaboration.org/artman2/uploads/1/2/D_Wilson__Calculating_ES.pdf

<http://dx.doi.org/10.4135/978144627305013510256>