Abstract Title Page
Not included in page count.

Title:

Which interventions have the greatest effect on student learning in Sub-Saharan Africa? *A meta-analysis of rigorous impact evaluations*

Authors and Affiliations:

Katharine Conn
PhD candidate, Economics and Education,
Columbia University, Teachers College

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

In the last three decades, there has been a large increase in the number of rigorous experimental and quasi-experimental evaluations of education programs in developing countries. These impact evaluations have taken place all over the globe, including a large number in Sub-Saharan Africa (SSA). For example, Blimpo and Evans (2011) examine the impact of school-based management training on subsequent student achievement in The Gambia, Piper and Korda (2011) estimate the impact of adaptive instruction & teacher coaching on student learning in Liberia, and Kremer, Miguel and Thornton (2009) examine the impact of merit scholarships on student achievement in Kenya.

Given the large number of programs in SSA, an important question is "what works" in terms of improving learning outcomes. Previous reviews have attempted to synthesize the large body of literature in developing countries, though not specific to SSA. Both Kremer and Holla (2009) and Kremer, Brannen, and Glennerster (2013) reviewed a selection of randomized controlled trials in the developing world and found that technology-assisted learning, remedial education, student tracking and the use of contract teachers were among the most effective interventions at improving student achievement. They note that technology-assisted learning and remedial education are particularly effective when the intervention is adaptive to the student's learning level but that inputs such as hiring additional teachers, buying more textbooks, or providing flexible school grants have surprisingly small effects. Further, McEwan (2013) conducted a meta-analysis of 76 randomized experiments from the developing world and found that intervention types with the largest average effect sizes were those that employed computer-assisted learning, teacher training, and smaller classes/smaller learning groups/ ability grouping. Intervention types that were least effective included school monetary grants, and nutritional/ health treatments.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

These previous syntheses, however, have focused broadly on the developing world, not specifically on interventions and programs in SSA. The fact that the developing world is socially and economically diverse and that different interventions and programs are often culturally specific suggests that the broad results from previous syntheses may not be applicable to SSA, which is itself a quite diverse region. In this paper, I conduct a large systematic review of the impact evaluation literature throughout the region. This includes both experimental and quasi-experimental studies from all fields which evaluate the impact of education programs on student learning (fields include economics, health, psychology, and education).

The purpose of this project is to better understand both which programs have been implemented and evaluated in the region and, of these, to determine which have the most promise in terms of increasing student learning outcomes in SSA – a critical policy imperative for developing countries with limited resources to spend. For example, do financial incentives for teachers result in similar outcomes as providing incentives to students or their families? Or are interventions that provide school supplies to students (textbooks, flipcharts) as effective as those that change the method of instruction? This paper focuses on analyzing this new body of rigorous education research in the context of SSA.

I also examine whether program effects differ by population subgroup (gender, income level etc.), region, education level (primary or secondary), subject matter (math, language, etc.), assessment type (researcher designed or standardized), target of intervention (student, classroom, teacher, or community), or intervention dosage (length or frequency of the intervention), among others. Lastly, I explore whether there are heterogeneous effects by study methodology (i.e., whether effects sizes are correlated with the type of experimental/ quasi-experimental strategy), study publication type, or study quality (quality index defined in full paper).

Setting:

Description of the research location.

In this paper, I focus on educational interventions aimed at improving learning outcomes in SSA. This region includes East, Central, Southern, and Western Africa, as well as the Indian Ocean Islands.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

The study population for this meta-analysis includes evaluations of education programs at the primary and secondary level. I include interventions aimed at all students and at particular subgroups of students (e.g., single-sex education). Additionally, I focus only on studies that have taken place since 1980.

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.

The focus of this paper is to better understand the effects of interventions aimed at improving learning outcomes. This includes over thirteen different intervention types, ranging from pedagogical interventions (i.e. cooperative learning strategies or bilingual medium of instruction), to interventions in school health, to programs that alter student or teacher incentives. Table 1 includes a full list and explanation of program types (please insert Table 1 here).

Within each intervention area, programs differ by treatment amount (treatments have different intensities or run for different amounts of time), the implementing agents of the program (NGOs, ministry of education, outside researchers), or by the specific sub-population targeted. Additionally, in this paper I focus only on experimental and quasi-experimental studies. This includes both large-scale, multi-site evaluations and single-site evaluations, as well as studies using propensity score matching, regression discontinuity, difference-in-differences, or instrumental variables designs. I exclude correlational studies.

Research Design:

Description of the research design.

For my research design, I employ a random effects meta-analysis model (Lipsey and Wilson, 2001). I chose a random effects model (as opposed to a fixed effects model) because the interventions under study are relatively disparate in their program characteristics, even within an intervention area (for example, within school health interventions, I examine both iodine provision as well as de-worming programs). Thus I employ a random-effects model which allows me to estimate mean effect sizes, as opposed to a “common” or “true” effect sizes (used with fixed effects models).

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

In order to better understand the literature, the first step of this systematic review was to conduct a large literature search. During this stage, careful attention was paid to the specification of inclusion/exclusion criteria as well as search terms. Efforts were made to include the gray literature – i.e., unpublished studies, such as those found through the World Bank and the MIT Poverty Action Lab, as well as other organizational publications.

This literature search resulted in over 10,660 articles. The inclusion/exclusion criteria were set based upon the population, setting, and interventions mentioned above. After reading abstracts and papers, the final analyses includes 60 primary studies (within which are 75 experiments). Table 2 indicates the process through which these 60 studies were selected (please insert Table 2 here). Importantly, one finding is that the number of studies taking place in SSA is much larger than that of previous reviews, which contained between 8 – 19 studies each (Kremer and Holla, 2009; Kremer, Brannen, and Glennerster, 2013; and McEwan, 2013).

The final set of studies was coded based on an extensive set of questions regarding the population, setting, intervention, and study design. In order to reduce researcher-bias, a random subset was coded by a second reviewer; the inter-rater reliability was just over 98%. Next, the outcomes were converted into effect sizes. In this paper, I focus on Cohen's *d*, the standardized mean difference, since this is the effect size commonly used in experiments. In many studies, particularly those evaluating pedagogical interventions, this involved correcting outcomes for clustering (Hedges, 2007). In order to combine and analyze these results, I then employ a random effects meta-analysis model (described above). Since many studies report multiple outcomes (without information on the correlation structure), in some analyses I use robust variance estimation (Hedges, Tipton, and Johnson, 2010), and when the number of studies is small, I use small-sample corrections (Tipton, in press). Finally, in addition to an overall analysis, I also conduct a "mini-meta-analysis" for each education topic (where data permits).

Findings / Results:

Description of the main findings with specific details.

Overall effect size: The overall pooled effect size across all interventions is 0.16 (SE=0.031; p = 0.0001). Importantly, the proportion of the observed variance in effect size estimates attributable to real difference across studies (I^2 statistic) is 80.1% ($\tau^2 = 0.024$). This I^2 value is quite high (and relatively rare), but not unexpected given that I am pooling evaluations of studies across many intervention types. The forest plot in Figure 1 displays the full set of effect sizes under study (please insert Figure 1 here).

Effect sizes by intervention type: Among those intervention types with the lowest pooled effect sizes are the abolishment of school fees (Cohen's d = 0.007, p = 0.484), school-based management interventions (d= 0.047, p = 0.278), and school supplies provision (d=0.058, p = 0.004). Interventions with the strongest the impact are comprehensive equity interventions (interventions that provide multiple inputs such as school infrastructure improvements, school meals, textbooks, and community programs as one package) [d = 0.38, p = 0.000]; instructional time increases (i.e. interventions that extend the school day) [d = 0.451, p = 0.000], and pedagogical interventions (i.e. interventions that implement new instructional methods) [d=1.128, p = 0.000]. The forest plot in Figure 2 displays the pooled effect sizes by intervention type (please insert Figure 2 here).

Sub-group/moderator effects: While I conduct numerous sub-group/moderator analyses in my full paper, due to space constraints, I will briefly discuss my results for only two analyses

here: 1.) Pedagogical interventions have an average effect size 0.862 standard deviations higher than any other intervention type (SE= 0.105, p-value = 0.0001). Further, I find that pedagogical interventions that focus on procedural learning (executing actions sequences to solve problems) or conceptual learning (learning through classification/from examples) have the highest pooled effect sizes of 1.830 (p = 0.000) and 2.075 (p = 0.008) standard deviations, respectively. Inquiry-based learning interventions with a pooled effect size of 0.257 (p = 0.330) are the least effective of the pedagogical interventions, but even these programs seem to have a reasonable impact when compared to all non-pedagogical interventions. The forest plot in Figure 3 displays the results of this "mini-meta-analysis." (please insert Figure 3 here).

2.) In my second sub-group analysis, I examine how the impact of an intervention may vary depending on the "target" of the intervention (i.e. whether the subject of the intervention was the student, classroom, teacher, or community). I find that education interventions that target the classroom (i.e. tracking, class size, or instructional interventions) have the highest average effect size followed closely by those that target students (i.e. merit scholarships, or provision of uniforms, and textbooks), then followed by those that target teachers (i.e. teacher monetary incentives or contractual teachers). Interventions that target the entire community (provision of school report cards to the community or school management committee trainings) are estimated to have the lowest average impact. See the forest plot in Figure 4 for the various effect size magnitudes (please insert Figure 4 here). Overall, I find that interventions that target students or the entire classroom to be 1.13 standard deviations (SE= 0.254, p = 0.001) higher than those that target the teacher or school community.

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

Overall, this meta-analysis finds that interventions in pedagogical methods (particularly those that emphasize procedural or conceptual learning) to be very promising. The high-magnitude effect sizes of these studies stand in contrast to interventions that have abolished school fees or improved school-based management (those with the lowest impact). This underscores the fact that interventions that may improve student enrollment or attendance (such as free/subsidized admission or community involvement) do not necessarily result in increased test scores; while children may attend school more frequently, they are not necessarily learning if school quality is low. However, evaluations of these high-impact pedagogical interventions are often quite small (randomization units under 30), and thus may suffer from various biases. Further, while other interventions that target the classroom (class size or student tracking) also seem to be promising, they are few in number.

In general, while there is more rigorous evidence on interventions in education than previously reported, this evidence is not evenly distributed across topics or countries. In fact, when performing my literature search, I found that certain topics that are very relevant to the African context (such as multi-grade teaching or multi-shift teaching) are not under rigorous study at all. Further, over 30% of the studies in this meta-analysis come from Kenya alone. These findings caution against a strict interpretation of these results, as program effects may differ by country as well within country, depending on the initial conditions and binding constraints of different educational systems. [Note: In my full paper, I do test for contextual differences in program impact]. Recommendations for researchers would thus include the use of larger randomized trials, particularly in the field of pedagogical research, and an increased focus on evaluations that are more representative of both the continent and the issues facing education systems across Africa.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons Ltd.

Cooper, H. M., Hedges, L. V., & Valentine, J. (Eds.) (2009). *The Handbook of Research Synthesis and Meta-Analysis (2nd Edition).* New York: The Russell Sage Foundation.

Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and Behavioral Statistics. 32* (4), 341–370.

Hedges, L.V., Tipton, E. & Johnson, M.C. (2010, January/March). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1* (1), 39–65.

Kremer, M. & Holla, A. (2009). Improving Education in the Developing World: What Have We Learned from Randomized Evaluations? *Annual Review of Economics, 1,* 513–542.

Kremer, M., Brannen, C. & Glennerster, R. (2013). The Challenge of Education and Learning in the Developing World. *Science, 340,* 297-299.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, Calif: Sage Publications.

McEwan, P. J. (2013, August). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. Unpublished Manuscript.

Tipton, E. (*in press*). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods.*

Appendix B. Tables and Figures
Not included in page count.

Table 1

Intervention Area:	Impact Evaluations examine:
Abolishment of School Fees	Universalization of schooling/ elimination of fees
Class Size	Class size reduction/increase
Comprehensive Equity Program	Programs which address inequalities in education through a myriad of strategies (for example, a combination of teacher training, supplemental education & increased instructional time)
Cash Transfers	Programs which offer monetary support to families, (can be conditional on their children's school enrollment/ attendance/performance)
Information for Accountability	Interventions that provide student/ school performance information to parents/ communities in order to increase transparency & accountability.
Instructional Time	Increase/decrease in the number of hours that students attend school (can be altered through the length of a school day, length of the school year, multi-grade teaching, multi-shift teaching).
Pedagogical Interventions	Programs that affect the method(s) of instruction & learning; Changes in language of instruction policies; ICT computer programs employed both in and out of the classroom.
School-Based Management	Devolution of certain powers/responsibilities regarding education provision to the local level (parents/teachers/school community)
School Health	Programs which include both nutritional (school meals) and other health-related interventions (vitamins, de-worming drugs etc.)
School Supplies Provision	Provision of flipcharts, textbooks, writing materials etc.
Student Incentives	The provision of scholarships (or other incentives such as gifts) to students in order to incentivize sustained enrollment or improved test scores.
Teacher Incentives	Teacher or school-wide bonus pay (based on student achievement, as well as other select variables); Provision of temporary teacher contracts.
Teacher Training	Training of teachers in both pedagogical methods as well as academic content areas.
Tracking & Peer Effects	Grouping of students by ability level; influence of students from certain backgrounds/ability levels on rest of student population.

Table 2:

Total relevant articles after following the search process above *(+ citation tracking etc.)*	**10,660**
Total articles meeting criteria (based on abstract alone)	**168**
After a more thorough read of these 168 studies:	
➢ No. w/o a rigorous methodology	-24
➢ No. w/o achievement outcomes	-29
➢ No. outside of SSA	-2
➢ No. w/o explict findings	-6
➢ No. duplicates (earlier dates)	-16
➢ No. w/ non-school-aged population	-3
➢ No. whose findings can't be standardized	-28
Total number of studies to be included in learning outcomes meta-analysis	**60**

Figure 1:

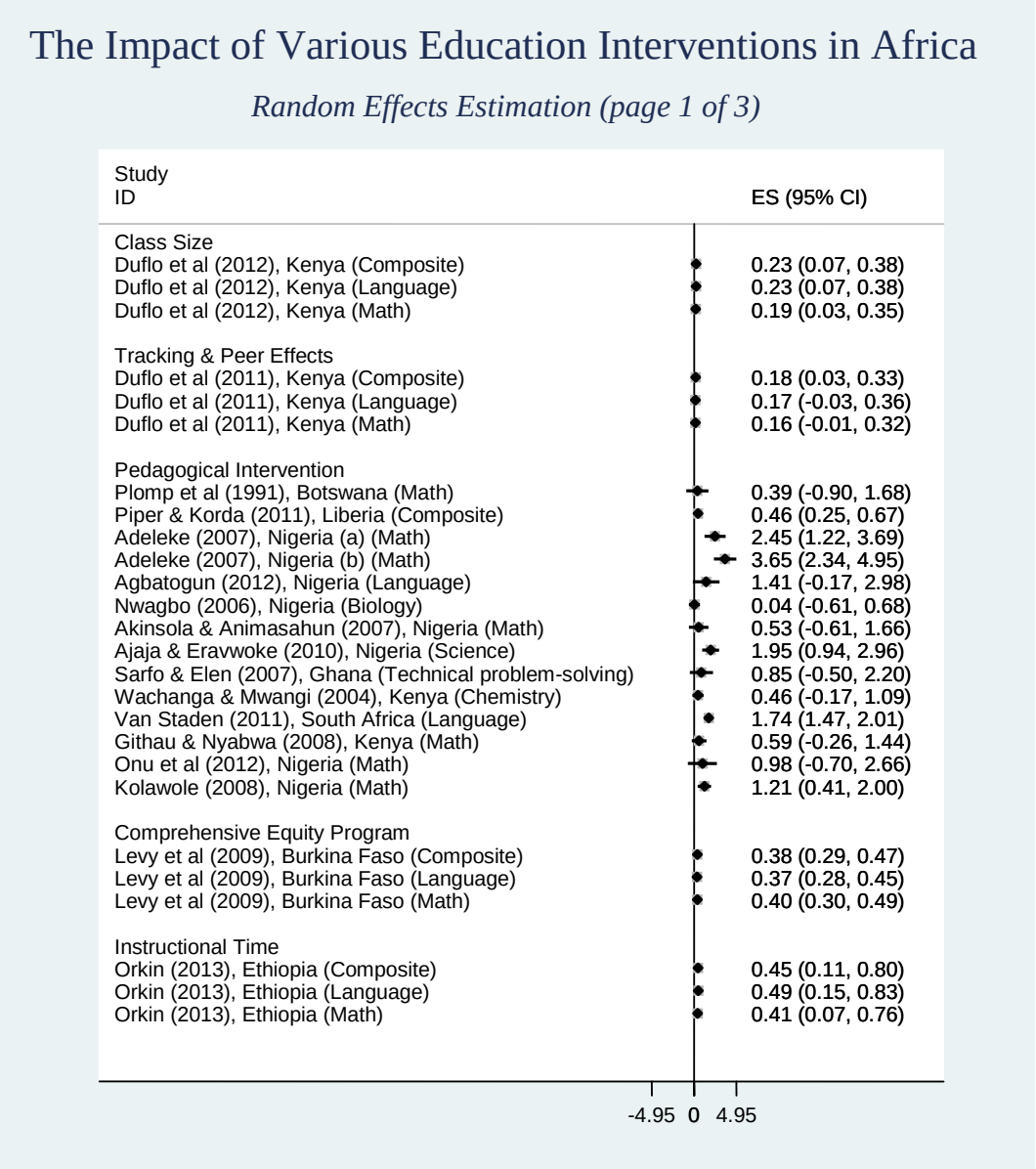

The Impact of Various Education Interventions in Africa
Random Effects Estimation (page 1 of 3)
Study ID
ES (95% CI)
Class Size
Duflo et al (2012), Kenya (Composite) 0.23 (0.07, 0.38)
Duflo et al (2012), Kenya (Language) 0.23 (0.07, 0.38)
Duflo et al (2012), Kenya (Math) 0.19 (0.03, 0.35)
Tracking & Peer Effects
Duflo et al (2011), Kenya (Composite) 0.18 (0.03, 0.33)
Duflo et al (2011), Kenya (Language) 0.17 (-0.03, 0.36)
Duflo et al (2011), Kenya (Math) 0.16 (-0.01, 0.32)
Pedagogical Intervention
Plomp et al (1991), Botswana (Math) 0.39 (-0.90, 1.68)
Piper & Korda (2011), Liberia (Composite) 0.46 (0.25, 0.67)
Adeleke (2007), Nigeria (a) (Math) 2.45 (1.22, 3.69)
Adeleke (2007), Nigeria (b) (Math) 3.65 (2.34, 4.95)
Agbatogun (2012), Nigeria (Language) 1.41 (-0.17, 2.98)
Nwagbo (2006), Nigeria (Biology) 0.04 (-0.61, 0.68)
Akinsola & Animasahun (2007), Nigeria (Math) 0.53 (-0.61, 1.66)
Ajaja & Eravwoke (2010), Nigeria (Science) 1.95 (0.94, 2.96)
Sarfo & Elen (2007), Ghana (Technical problem-solving) 0.85 (-0.50, 2.20)
Wachanga & Mwangi (2004), Kenya (Chemistry) 0.46 (-0.17, 1.09)
Van Staden (2011), South Africa (Language) 1.74 (1.47, 2.01)
Githau & Nyabwa (2008), Kenya (Math) 0.59 (-0.26, 1.44)
Onu et al (2012), Nigeria (Math) 0.98 (-0.70, 2.66)
Kolawole (2008), Nigeria (Math) 1.21 (0.41, 2.00)
Comprehensive Equity Program
Levy et al (2009), Burkina Faso (Composite) 0.38 (0.29, 0.47)
Levy et al (2009), Burkina Faso (Language) 0.37 (0.28, 0.45)
Levy et al (2009), Burkina Faso (Math) 0.40 (0.30, 0.49)
Instructional Time
Orkin (2013), Ethiopia (Composite) 0.45 (0.11, 0.80)
Orkin (2013), Ethiopia (Language) 0.49 (0.15, 0.83)
Orkin (2013), Ethiopia (Math) 0.41 (0.07, 0.76)
-4.95
0
4.95

Figure 1 (continued):

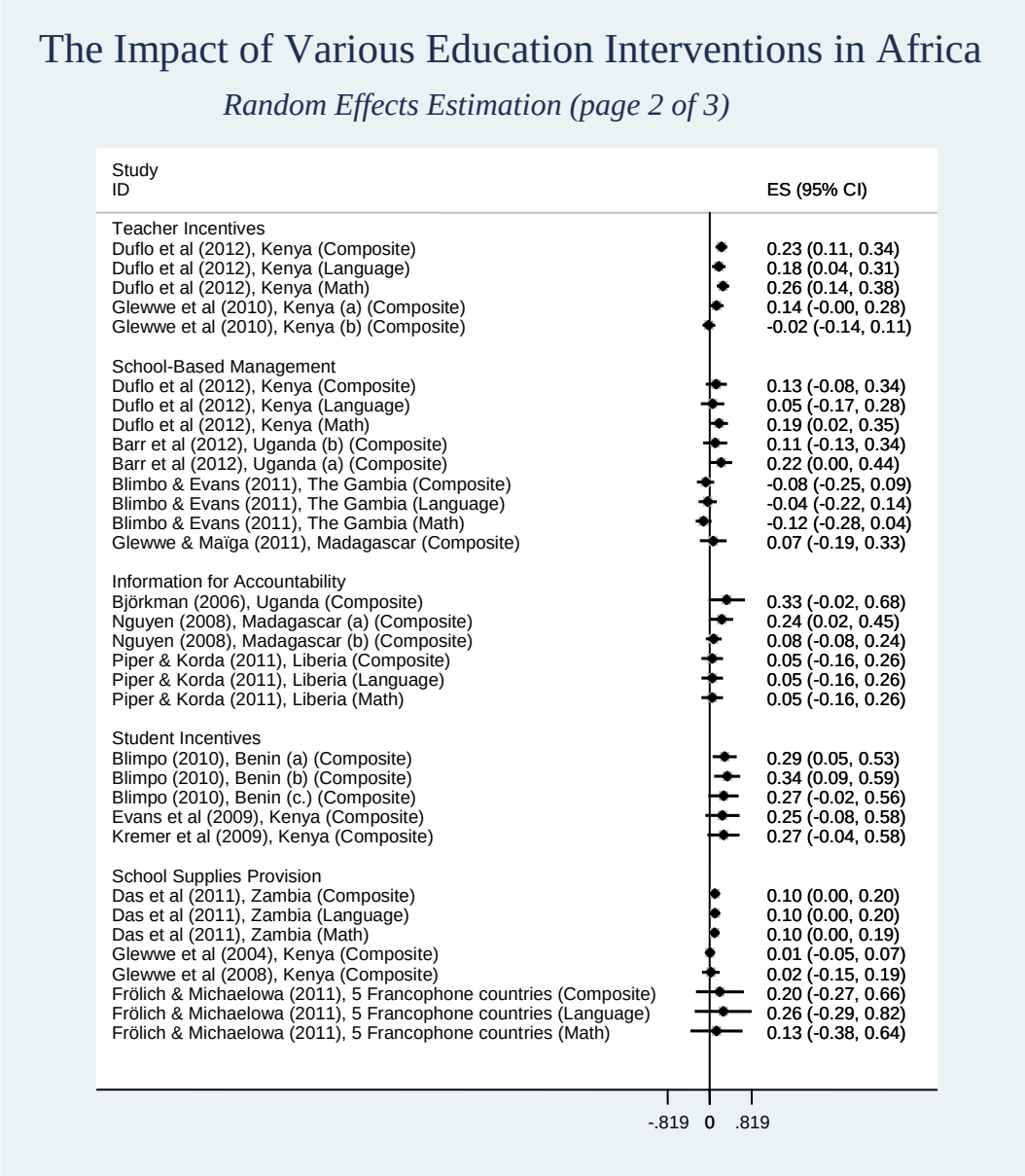
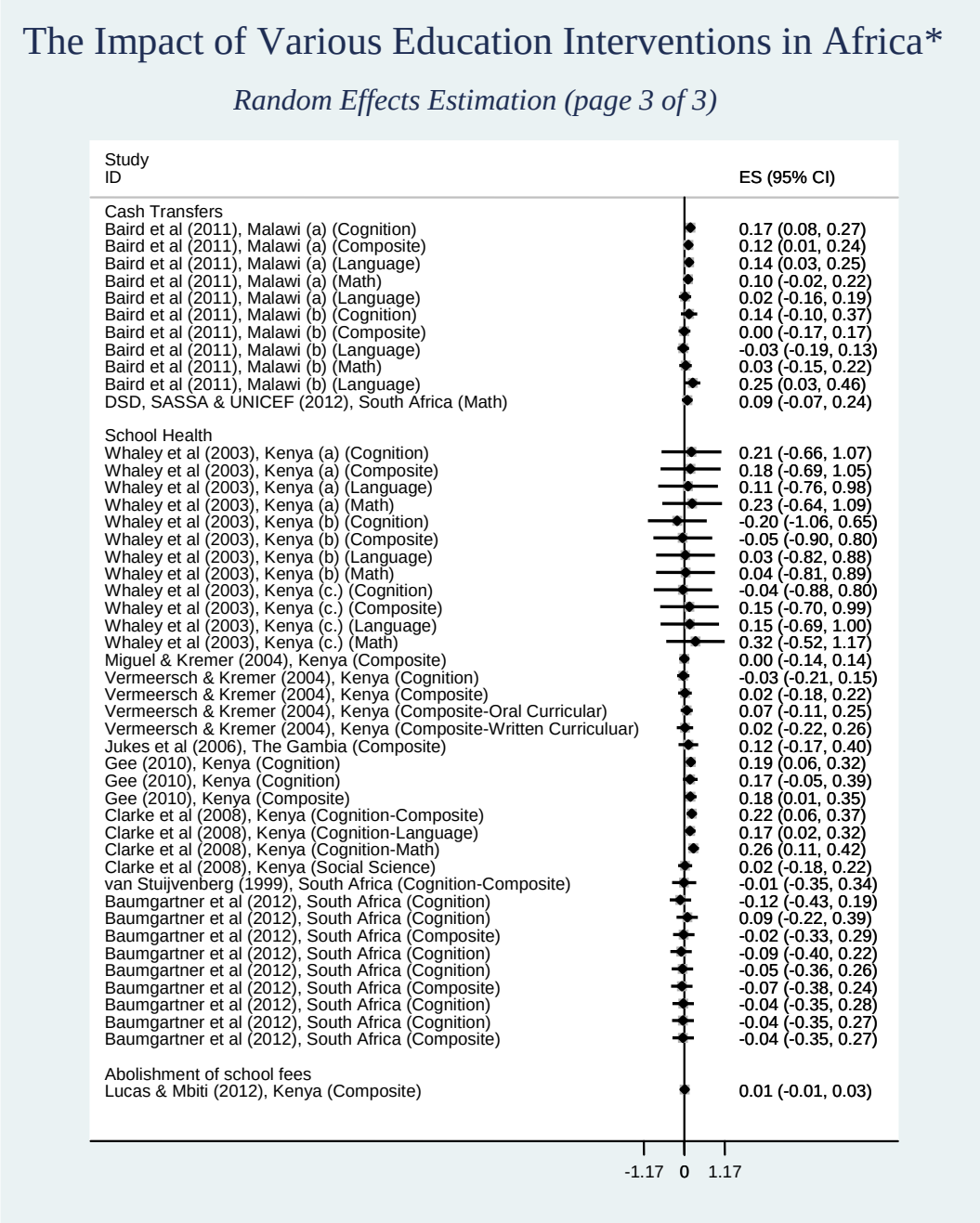


Figure 1 (continued):



*NOTE: The purpose of this Forest Plot is to illustrate the effect sizes and variances used in this meta-analysis. It does not include a total effect size, or weights, since code currently does not exist for integrating forest plots into the RVE framework. See the text for the overall average effect size.

Figure 2: Forest plot for random effects meta-analysis, all intervention types.

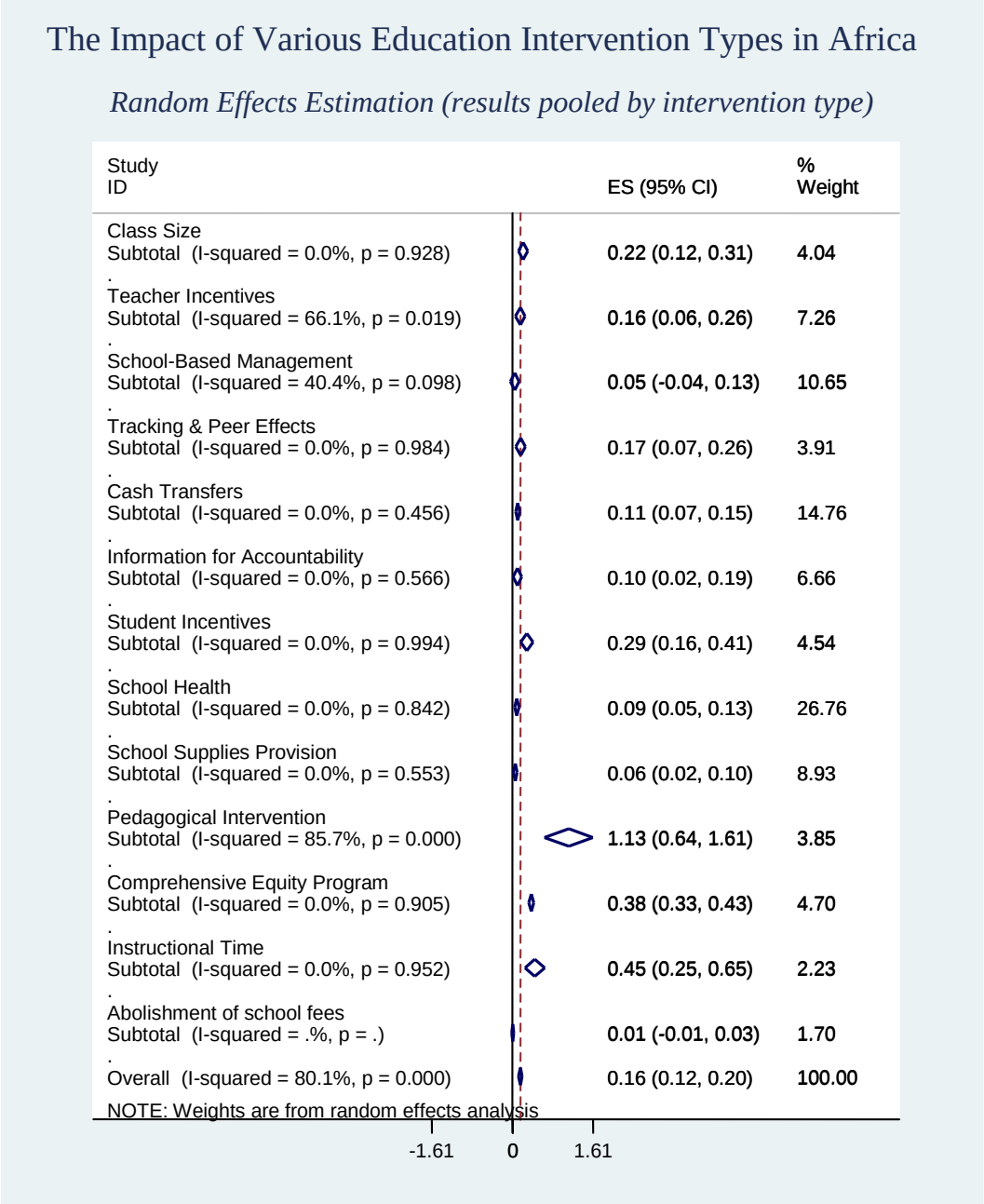


Figure 3: Forest plot for random effects meta-analysis, pedagogical interventions only.

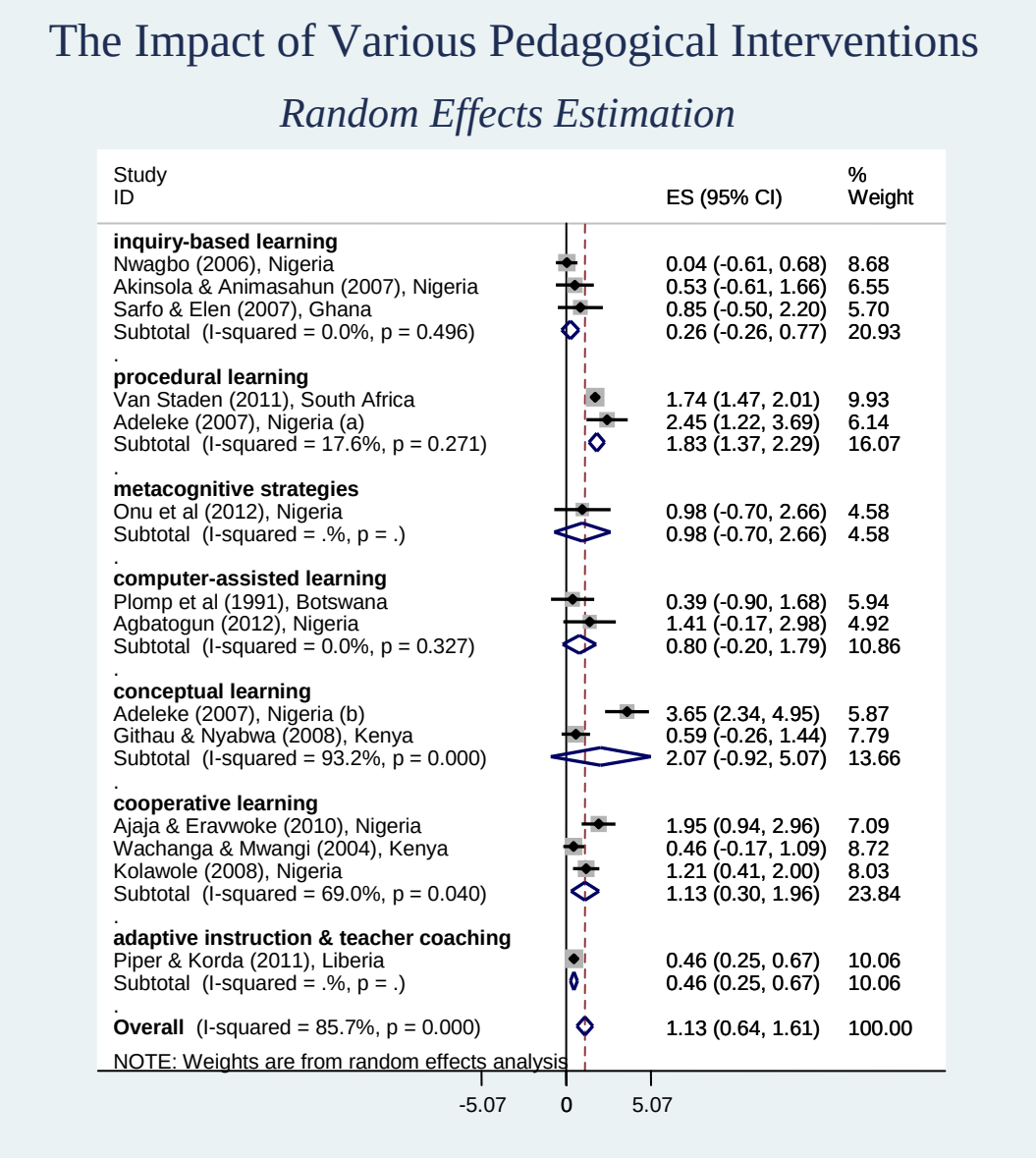


Figure 4: Forest plot for random effects meta-analysis (sub-group = intervention target)