How Big Are Effect Sizes in International Education Studies?

David K. Evans and Fei Yuan

Abstract

A growing literature measures the impact of education interventions in low- and middle-income countries on both access and learning outcomes. But how should one contextualize the size of impacts? This article provides the distribution of standardized effect sizes on learning and access from 234 studies in low- and middle-income countries. We identify a median effect size of 0.10 standard deviations on learning and 0.07 standard deviations on access among randomized controlled trials. Effect sizes are similar for quasi-experimental studies. Effects are larger and demonstrate higher variance for small-scale studies than for large-scale studies. The distribution of existing effects can help researchers and policymakers to situate new findings within current knowledge and design new studies with sufficient statistical power to identify effects.

Keywords: achievement, assessment, international education/studies, literacy, effect size, econometric analysis

This working paper was originally published in August 2020. It was updated in April 2022. The original version can be viewed at https://www.cgdev.org/sites/default/files/2022-04/how-big-are-effect-sizes-international-education-studies.pdf.

CENTER FOR GLOBAL DEVELOPMENT

www.cgdev.org

Working Paper 545
April 2022

How Big Are Effect Sizes in International Education Studies?

David K. Evans
Center for Global Development
devans@cgdev.org

Fei Yuan
Harvard Graduate School of Education
fyuan@g.harvard.edu

Author order is alphabetical. The authors thank Lee Crawfurd, Susannah Hares, Matthew Kraft, Justin Sandefur, Eric Taylor, and Eva Vivalt for helpful comments and Amina Mendez Acosta for helpful research assistance.

This article was published in Educational Evaluation and Policy Analysis in 2022.

The Center for Global Development is grateful for contributions from the Bill & Melinda Gates Foundation in support of this work.

David K. Evans and Fei Yuan. 2022. "How Big Are Effect Sizes in International Education Studies?." CGD Working Paper 545. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/how-big-are-effect-sizes-international-education-studies

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Briefs

Educational Evaluation and Policy Analysis
Month 202X, Vol. XX, No. X, pp. 1–9
DOI: 10.3102/01623737221079646
Article reuse guidelines: sagepub.com/journals-permissions

How Big Are Effect Sizes in International Education Studies?

David K. Evans
Center for Global Development

Fei Yuan
Harvard Graduate School of Education

A growing literature measures the impact of education interventions in low- and middle-income countries on both access and learning outcomes. But how should one contextualize the size of impacts? This article provides the distribution of standardized effect sizes on learning and access from 234 studies in low- and middle-income countries. We identify a median effect size of 0.10 standard deviations on learning and 0.07 standard deviations on access among randomized controlled trials. Effect sizes are similar for quasi-experimental studies. Effects are larger and demonstrate higher variance for small-scale studies than for large-scale studies. The distribution of existing effects can help researchers and policymakers to situate new findings within current knowledge and design new studies with sufficient statistical power to identify effects.

Keywords: *achievement, assessment, international education/studies, literacy, effect size, econometric analysis*

Introduction

Recent years have seen a rapid increase in rigorous evaluations of interventions intended to expand educational access and improve learning outcomes (World Bank, 2018). But how big is a big effect size? Researchers need an accurate sense of expected effect sizes to design future studies, and policymakers need to know what kinds of effect sizes have been identified in the past to anchor expectations for the future. Many researchers draw on a rule of thumb proposed by Cohen (1969) that a small effect size is 0.2 standard deviations (*SD*s) and a large effect size is 0.8 *SD*s (Becker & Park, 2011; Klassen & Tze, 2014; Piper & Mugenda, 2014). More recently, Kraft (2020) examined the distribution of more than 700 studies with learning outcomes in high-income countries and found a median impact of just 0.1 *SD*; even the 90th percentile effect size was under 0.5 *SD*s, suggesting that expectations based on old benchmarks may be too high.

In this article, we present the distribution of effect sizes from 234 studies in low- and middle-income countries that evaluate the impact of educational interventions on students' access to schooling or on students' learning in school. We further provide the distribution of effect sizes across different study designs (randomized controlled trials [RCTs] vs. quasi-experimental studies), the scale of the study, and the specific learning outcome (e.g., math vs. reading). This distribution of effects provides researchers with a simple way to situate impacts of new studies relative to what is known about how to expand access and increase learning. It also anchors expectations about the impact of future interventions. The current distribution of effect sizes does not rule out dramatically larger effect sizes of innovative education interventions in the future, but it does help researchers and policymakers to understand what a dramatically larger effect size would be.

Specifically, we draw on a large database of 138 RCTs and 96 quasi-experimental studies with learning or access outcomes, primarily obtained from previous reviews of education in low- and middle-income countries. We then standardize effect sizes across studies. We find that across the 96 of 138 RCTs that report learning outcomes (math or reading), the median effect size is 0.10 *SD*s, with a 25th percentile of 0.01, a 75th percentile of 0.23, and a 90th percentile—to give the reader a sense of what a large effect size looks like—of 0.45 *SD*s. (In his work in high-income countries, Kraft, 2020, also identifies a median impact of 0.10 *SD*s across education interventions.) Studies with a sample size in the bottom quartile report point estimates that are on average almost 3 times the size of studies with a sample size in the top quartile: 0.27 *SD*s versus 0.10 *SD*s. For access outcomes (i.e., enrollment, attendance, and dropout), the median effect size is 0.07 *SD*s across 73 RCTs, with a 25th percentile of 0.02, a 75th percentile of 0.14, and 90th percentile of 0.30 *SD*s. The distribution of access effect sizes is similar whether the outcome is enrollment or attendance. Our sample of quasi-experimental studies shows a similar pattern of results.

This distribution of effects contrasts with the benchmarks proposed by Cohen (1969). Those effect sizes were developed based on a small sample of social psychology lab experiments in the United States in the 1960s, mostly with undergraduate students (Kraft, 2020), so their relevance to impact evaluations in basic education today is questionable, even more so in low- and middle-income countries.

This study contributes to the growing body of synthesis work on the impact of education interventions in low- and middle-income countries (Evans & Popova, 2016; Kremer et al., 2013; Snilstveit et al., 2015), as well as to work seeking to characterize how large the impact of education interventions is relative to benchmarks such as the amount of learning usually gained during a year of schooling (Evans & Yuan, 2019) or the difference in learning levels between rich and poor countries (Angrist et al., 2020; Filmer et al., 2020).

Data and Analysis

We constructed a database of education impact evaluation studies to collect effect sizes. First, we draw on all studies included in 11 existing systematic reviews of education interventions in low- and middle-income countries. Then, we carry out a limited search to update the evidence that came out after those reviews were published, through early 2018. We detail our search, selection, and standardization processes in Supplemental Appendix A in the online version of the journal.

In brief, we include experimental and quasi-experimental evaluations of three types of programs. We include direct education interventions (such as teacher professional development or providing learning materials), and we include two other classes of interventions that commonly report educational outcomes: health interventions (such as providing deworming drugs and micronutrients) and safety net interventions (such as cash transfers). Studies also need to report at least one access outcome (e.g., enrollment, dropout, or attendance) or a learning outcome (e.g., literacy or numeracy). We focus on programs that took place in preprimary, primary, and secondary schools.

Our final analytical sample consists of 234 studies with 942 effect sizes from 51 low- and middle-income countries between 1981 and late 2017. Table 1 presents the summary statistics of our sample. Although our sample ends in late 2017, we find that effect sizes have remained consistent since 2006 (see Supplemental Appendix Figure B1 in the online version of the journal) suggesting that more recent studies are unlikely to be different.

Our sample includes 683 effect sizes from 138 RCTs and 259 effect sizes from 96 quasi-experimental studies. We categorized measured outcomes into two broad types: learning (which includes test scores of reading and math) and access (which includes enrollment, attendance, and dropout). We report the absolute value of estimates of impacts on dropout rates, so that positive numbers are always associated with

TABLE 1

Summary Statistics

	All	RCTs	Quasi-experimental studies
Number of all studies	234	138	96
Number of effect sizes	942	683	259
Number of countries	51	39	38
Learning outcomes			
Number of effect sizes	600	468	132
Reading	348	269	79
Math	252	199	53
No. of studies with learning outcomes	139	96	43
Access outcomes			
Number of access effect sizes	342	215	127
Enrollment	162	76	86
Attendance	130	105	25
Dropout	50	34	16
No. of studies with access outcomes	137	73	64
By region (number of studies)[a]			
Sub-Saharan Africa	81	53	28
South Asia	53	34	19
Latin America and the Caribbean	89	39	50
East Asia and Pacific	51	36	15
Europe and Central Asia	3	0	3
Middle East and North Africa	1	1	0
By region (effect sizes)			
Sub-Saharan Africa	334	263	71
South Asia	210	144	66
Latin America and the Caribbean	237	143	94
East Asia and Pacific	148	124	24
Europe and Central Asia	4	0	4
Middle East and North Africa	9	9	0

Source. Authors' construction.
Note. Region groups follow World Bank (2020). RCTs = randomized controlled trials.
[a]One study (Das et al., 2013) includes countries in both South Asia and Sub-Saharan Africa.

improvements. Sub-Saharan Africa has the largest number of effect sizes (334), whereas Europe and Central Africa (4) and Middle East and North Africa (9) have the fewest.

Results

Across the RCTs that measure reading or math outcomes, we find a median impact of 0.10 *SD*s (Table 2). The median impact is smaller for math assessments (0.07 *SD*s) than for reading assessments (0.14 *SD*s). For small studies with a sample size in the bottom quartile (under 732 participating students), median impacts are 0.23 *SD*s. For large studies, with a sample size in the top quartile (with more than 4,974 students), the median impact is a fourth of that, at 0.06 *SD*s. The variation is also higher among smaller studies. The distribution of impacts is comparable (if slightly smaller) for quasi-experimental studies (see Supplemental Appendix Table B1, Panel A in the online version of the journal), including the finding of larger impacts for the smallest scale studies and smaller impacts for the largest scale studies.

TABLE 2

Distribution of Learning Impacts Across RCTs

		Subject		Sample size			
	Overall	Reading	Math	First quartile (≤732)	Second quartile (732, 2,048)	Third quartile (2,048, 4,974)	Fourth quartile (>4,974)
M	0.16	0.19	0.11	0.27	0.18	0.11	0.10
SD	0.25	0.23	0.27	0.39	0.26	0.15	0.15
P1	−0.22	−0.22	−0.36	−0.14	−0.49	−0.21	−0.18
P10	−0.04	−0.03	−0.05	−0.04	−0.12	−0.05	−0.01
P20	0.01	0.01	−0.01	0.03	−0.02	0.00	0.01
P25	0.01	0.03	0.01	0.06	0.02	0.01	0.01
P30	0.03	0.05	0.02	0.09	0.04	0.03	0.01
P40	0.06	0.09	0.05	0.15	0.10	0.06	0.04
P50	0.10	0.14	0.07	0.23	0.14	0.09	0.06
P60	0.14	0.18	0.10	0.28	0.18	0.12	0.08
P70	0.18	0.25	0.13	0.35	0.28	0.15	0.12
P75	0.23	0.32	0.15	0.37	0.34	0.17	0.13
P80	0.29	0.35	0.17	0.44	0.36	0.20	0.14
P90	0.45	0.50	0.31	0.56	0.49	0.32	0.24
P99	0.88	0.90	0.80	3.03	0.92	0.69	0.70
No. of effect sizes	468	269	199	81	133	122	132
No. of studies	96	70	75	20	33	32	27

Source. Authors' construction.
Note. Overall includes only reading and math effect sizes. The distribution of sample size is based on all RCT studies, not just those with reading or math outcomes; reported effect sizes by sample size quartile only include reading and math outcomes. RCT = randomized controlled trial.

For RCTs that report impacts on access, the median impact is smaller: 0.07 *SD*s (Table 3). Studies commonly report one or more of three access outcomes: enrollment (0.06 *SD*s), attendance (0.08 *SD*s), and dropout (0.05 *SD*s). The differences across outcomes are modest, but the slightly higher median for attendance may reflect greater ease in boosting student participation at the intensive margin than the extensive margin. The gap between small-scale and large-scale studies is similarly striking with access outcomes: For studies with a sample size in the bottom quartile, the median impact is 0.16 *SD*s, whereas for the largest studies (with a sample size in the top quartile), the median impact is just 0.05 *SD*s. Quasi-experimental studies again show slightly smaller effects and a similar pattern vis-à-vis the scale of the program (see Supplemental Appendix Table B1, Panel B in the online version of the journal).

Standard deviations are not always comparable across studies. As Singh (2015a, 2015b) shows, standard deviations will vary both across populations and across classes of tests. Thus, comparing standard deviations across contexts and tests should be treated with caution. In this study, we divide tests by subject to increase comparability. (In our main results, we only include reading and math effects. Supplemental Appendix Table B2, in the online version of the journal, reports the distribution of effect sizes for other types of tests—for example, composite tests—although a lack of comparability across those tests makes those results more difficult to interpret.) Furthermore, tests vary by other characteristics: They may be administered orally or in written format, administered in a group or individually, multiple choice or open-ended, or focused on one specific domain of—for example—reading or mathematics (see Supplemental Appendix Table B3 in the online version of the journal). We find that test score

TABLE 3

Distribution of Access Impacts Across RCT Trials

		Subject			Sample size			
	Overall	Enrollment	Attendance	Dropout (Abs. value)	First quartile (≤732)	Second quartile (732, 2,048)	Third quartile (2,048, 4,974)	Fourth quartile (>4,974)
M	0.10	0.11	0.12	0.05	0.22	0.17	0.11	0.09
SD	0.14	0.16	0.13	0.09	0.35	0.25	0.16	0.14
P1	−0.10	−0.10	−0.09	−0.12	−0.14	−0.43	−0.40	−0.16
P10	−0.02	−0.03	0.00	−0.07	−0.04	−0.09	−0.05	−0.02
P20	0.01	0.00	0.01	−0.02	0.01	−0.02	0.00	0.00
P25	0.02	0.02	0.02	0.00	0.04	0.03	0.01	0.01
P30	0.03	0.03	0.03	0.02	0.05	0.05	0.03	0.01
P40	0.05	0.05	0.07	0.03	0.10	0.10	0.06	0.03
P50	0.07	0.06	0.08	0.05	0.16	0.14	0.09	0.05
P60	0.09	0.08	0.11	0.06	0.23	0.17	0.13	0.08
P70	0.12	0.11	0.14	0.09	0.28	0.26	0.15	0.12
P75	0.14	0.13	0.15	0.10	0.33	0.30	0.18	0.13
P80	0.16	0.17	0.17	0.11	0.36	0.35	0.20	0.15
P90	0.30	0.38	0.35	0.17	0.55	0.47	0.28	0.23
P99	0.61	0.80	0.58	0.30	2.75	0.92	0.66	0.70
No. of effect sizes	215	76	105	34	112	164	165	158
No. of studies	73	33	44	16	34	40	45	38

Source. Authors' construction.
Note. Overall includes enrollment, attendance, and dropout (absolute value) effect sizes. The distribution of sample size is based on all RCT studies, not just those with enrollment, access or dropout (absolute value) outcomes; reported effect sizes by sample size quartile only include enrollment, access, or dropout outcomes. RCT = randomized controlled trial.

impacts tend to be much higher for orally administered and individually administered tests; this may in part be a function of the administration format and may also be because tests of more basic skills, like letter recognition, tend to be administered to younger children and so require oral, one-on-one administration. Test score impacts are also double for multiple-choice tests than what they are for tests with open-ended questions.

We further examine the relationship between scale of intervention and effect size in the three countries in our sample with the most estimates from RCTs, to enhance comparability across populations: Kenya (117 estimates), India (115 estimates), and China (72 estimates). For access estimates, we observe a negative relationship for all three countries; it is statistically significant in Kenya and China (see Supplemental Appendix Figure B2 in the online version of the journal). For learning estimates, we observe a negative, statistically significant relationship for Kenya and China, with no clear correlation for India (see Supplemental Appendix Figure B3 in the online version of the journal). As reported earlier and in Tables 2 and 3, we also see higher variation in small studies (related to the phenomenon documented in Barrera-Osorio & Ganimian, 2016; Kane & Staiger, 2002). If there is stronger publication bias among smaller studies, with it being easier to publish significant positive impacts, then we could observe higher average effects for smaller studies simply because of publication bias. As a partial test for such publication bias, we compare the results in journal articles with those in other publications (working papers, conference papers, or other reports). We find broadly comparable effects (see Supplemental Appendix Table B4 in the online version of the journal): The median effect size for learning outcomes from RCTs in journals versus other publications is 0.09 versus 0.11; for quasi-experimental studies, the parallel numbers are 0.05 and 0.06. The pattern for access outcomes is similar. If publication bias does not drive these effects, then it may simply be that implementing pilots effectively is easier than implementing at-scale programs.

Finally, we examine whether there are apparent differences in the distribution of effect sizes across regions (see Supplemental Appendix Table B5 in the online version of the journal). We find, in the four regions with a reasonable sample of studies, both the most RCTs and largest median impacts on learning in Sub-Saharan Africa (0.13 *SD*s).

Discussion

In this article, we provide a distribution of effect sizes from studies that measure the impact of education interventions on learning and access in low- and middle-income countries. These data can help to situate future studies among the distribution of existing work. This is not a normative distribution. There is a large gap between student access and student learning in low-income countries versus high-income countries (Filmer et al., 2020), and one can reasonably argue that closing that gap will require either much larger effect sizes or a great many reforms that deliver effect sizes of the type that we observe.

Our finding that—among RCTs—effect sizes in the smallest studies are quadruple those in larger studies for learning and triple for access also encourages caution when policymakers encounter pilot results with impressive effect sizes. It is possible to improve both access and learning at scale, but usually the improvements are smaller than those observed in pilots. Alternatively, we cannot rule out that this effect is at least partly driven by selective publication bias.

Our analysis yields recommendations for researchers. First, because standard deviations are not always comparable across studies, benchmarking effect sizes against real-world metrics can enhance interpretation (Eble & Escueta, 2021). For example, a recent study of a public–private partnership for primary education in Liberia yielded an effect size of 0.16 *SD*s after 3 years, which the authors emphasize is "equivalent to 4 words per minute additional reading fluency" (Romero & Sandefur, 2019). An early child intervention in Kenya increased children's language and motor development by 0.1 *SD*s, which the authors highlight as the same as the difference between "children of primary-educated and secondary-educated mothers" (Jakiela et al., 2020). Second, many tests used in development studies are designed by the research team instead of using standardized tests, and little is reported about what exactly is measured. Developing comparable measures across studies will require greater reporting precision by researchers.

One limitation of this work is that we do not incorporate cost effectiveness, in part because a careful treatment of costs merits a full separate study of its own. Yet, the interpretation of an effect size is mediated by the cost of the intervention. Unfortunately, a minority of studies report cost effectiveness into their analysis and these studies may not be representative. An analysis of 76 RCTs in low- and middle-income countries found that nearly half reported no details on costs and most of the others had minimal information (McEwan, 2015). A more recent analysis of recent education research from Africa found that less than a third of studies reported cost effectiveness (Evans & Mendez Acosta, 2021). As more studies report cost data in comparable ways, future research may supplement this work with cost analysis.

Readers may be tempted to despair that impacts of interventions tested thus far have been so small. In response to that, first, we highlight that comparing the literacy skills of adults with different years of schooling in five low- and middle-income countries (Bolivia, Colombia, Ghana, Kenya, and Vietnam) suggests that students gain between 0.15 and 0.21 *SD*s of literacy during the course of a school year (Evans & Yuan, 2019). Using longitudinal data, Singh (2020) estimates larger impacts for a year of schooling, ranging from 0.35 to 0.40 *SD*s in Ethiopia and India to 0.75 *SD*s in Vietnam. Thus, a large effect on learning according to our distribution (e.g., larger than 0.23 *SD*s, or three quarters of the distribution) is the equivalent to how much a student might learn in more than a full year of business-as-usual schooling according to the Evans and

Yuan (2019) estimates or between 58% and 66% of a year of schooling in Ethiopia or India according to the Singh (2020) estimates. Either way, these are substantive impacts.

Second, average effects can mask large effects for subsets of students. An average effect of 0.23 *SD*s (again, a large effect size) may mean that an intervention had no impact for three quarters of the students but an impact of 0.92 *SD*s for one quarter of students, which would be considered a success by virtually any metric (Gelman, 2020). As practical examples, a conditional cash transfer program with educational conditions reduced dropout rates by twice as much for students who were performing poorly at baseline in China (Mo et al., 2013) and increased enrollment by nearly twice as much for younger children in Burkina Faso (Akresh et al., 2013).

Third, there are programs that do have exceptionally large program impacts. For example, a literacy program that provided intensive training and materials to teachers in Uganda to help them teach literacy in children's mother tongue increased reading scores by 0.64 *SD*s and writing scores by 0.45 *SD*s (Kerwin & Thornton, 2021). Scholarships to secondary school students in Ghana boosted enrollment by 0.56 *SD*s (Duflo et al., 2019). Providing take-home meals to students in Uganda, conditional on their attendance, boosted enrollment by 0.42 *SD*s (Alderman et al., 2012). These are all above the 90th percentile of the distribution of effect sizes. Interventions can and should aim to achieve large changes in learning and access, but it is important to understand the full distribution to understand the potential of existing educational interventions to close education gaps between high- and low-income countries.

Acknowledgments

The authors thank Lee Crawfurd, Susannah Hares, Matthew Kraft, Paco Martorell, Justin Sandefur, Eric Taylor, Eva Vivalt, and one anonymous referee for comments and Amina Mendez Acosta for research assistance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Bill & Melinda Gates Foundation provided financial support.

ORCID iD

David K. Evans https://orcid.org/0000-0001-5413-9284

References

Akresh, R., de Walque, D., & Kazianga, H. (2013). *Cash transfers and child schooling: Evidence from a randomized evaluation of the role of conditionality*. World Bank. https://doi.org/10.1596/1813-9450-6340

Alderman, H., Gilligan, D. O., & Lehrer, K. (2012). The impact of food for education programs on school participation in Northern Uganda. *Economic Development and Cultural Change*, *61*(1), 187–218. https://doi.org/10.1086/666949

Angrist, N., Evans, D. K., Filmer, D. P., Glennerster, R., Rogers, H., & Sabarwal, S. (2020). *A new micro measure for education interventions: Learning-Adjusted Years of Schooling (LAYS)* [Working Paper]. The World Bank.

Barrera-Osorio, F., & Ganimian, A. J. (2016). The barking dog that bites: Test score volatility and school rankings in Punjab, Pakistan. *International Journal of Educational Development*, *49*, 31–54. https://doi.org/10.1016/j.ijedudev.2016.01.004

Becker, K., & Park, K. (2011). Integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A meta-analysis. *Journal of STEM Education: Innovations and Research*, *12*(5), 23–37.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences* (1st ed.). Academic Press.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., & Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, *5*(2), 29–57.

Duflo, E., Dupas, P., & Kremer, M. (2019). *The impact of free secondary education: Experimental evidence from Ghana* [Massachusetts Institute of Technology Working Paper Cambridge]. https://web.stanford.edu/~pdupas/DDK_GhanaScholarships.pdf

Eble, A., & Escueta, M. (2021). *When bootstraps aren't enough: Aspirations, learning, and educational supply* (No. CDEP-CGEG WP No. 99). https://cdep.sipa.columbia.edu/sites/default/files/cdep/WP99Eble.pdf

Evans, D., & Mendez Acosta, A. (2021). Education in Africa: What are we learning? *Journal of African Economies*, *30*(1), 13–54. https://academic.oup.com/jae/article/30/1/13/5999001

Evans, D., & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*(2), 242–270. https://doi.org/10.1093/wbro/lkw004

Evans, D., & Yuan, F. (2019). *Equivalent years of schooling: A metric to communicate learning gains in concrete terms* (No. WPS8752; pp. 1–52). The World Bank. http://documents.worldbank.org/curated/en/123371-550594320297/Equivalent-Years-of-Schooling-A-Metric-to-Communicate-Learning-Gains-in-Concrete-Terms

Filmer, D., Rogers, H., Angrist, N., & Sabarwal, S. (2020). Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education. *Economics of Education Review*, *77*, 101971. https://doi.org/10.1016/j.econedurev.2020.101971

Gelman, A. (2020). *Understanding the "average treatment effect" number*. https://statmodeling.stat.columbia.edu/2020/06/30/understanding-the-average-treatment-effect-number/

Jakiela, P., Ozier, O., Fernald, L. C. H., & Knauer, H. A. (2020). *Big sisters* [Working paper]. http://www.pamjakiela.com/bigsisters.pdf

J-PAL. (2017). *Roll call: Getting children into school*. Policy Bulletin. https://www.povertyactionlab.org/sites/default/files/publication/roll-call-getting-children-into-school.pdf

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*(4), 91–114. https://doi.org/10.1257/089533002320950993

Kerwin, J. T., & Thornton, R. L. (2021). Making the grade: The sensitivity of education program effectiveness to input choices and outcome measures. *The Review of Economics and Statistics*, *103*(2), 251–264. https://doi.org/10.1162/rest_a_00911

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59–76.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X2091-2798

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, *340*(6130), 297–300. https://doi.org/10.1126/science.1235350

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, *85*(3), 353–394. https://doi.org/10.3102/0034654314553127

Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S., & Brinton, C. (2013). School dropouts and conditional cash transfers: Evidence from a randomised controlled trial in rural China's junior high schools. *The Journal of Development Studies*, *49*(2), 190–207. https://doi.org/10.1080/00220388.2012.724166

Piper, B., & Mugenda, A. (2014). *The Primary Math and Reading (PRIMR) initiative: Endline impact evaluation*. USAID. https://pdf.usaid.gov/pdf_docs/PA00K27S.pdf

Romero, M., & Sandefur, J. (2019). *Beyond short-term learning gains: The impact of outsourcing schools in Liberia after three years* (Center for Global Development Working Paper 521). https://www.cgdev.org/publication/beyond-short-term-learning-gains-impact-outsourcing-schools-liberia-after-three-years

Singh, A. (2015a, January 13). *How standard is a standard deviation? A cautionary note on using SDs to compare across impact evaluations in education* [Development Impact—World Bank Blog]. https://blogs.worldbank.org/impactevaluations/how-standard-standard-deviation-cautionary-note-using-sds-compare-across-impact-evaluations

Singh, A. (2015b). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics*, *113*, 16–32. https://doi.org/10.1016/j.jdeveco.2014.10.004

Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, *18*(4), 1770–1813. https://doi.org/10.1093/jeea/jvz033

Snilstveit, B., Stevenson, J., Philips, D., Vojtkova, M., Gallagher, E., Schimdt, T., Jobse, H., Geelen, M., Pastorello, M. G., & Eyers, J. (2015). *Interventions for improving learning outcomes and access to education in low- and middle-income countries: A systematic review* (3ie Systematic Review 24). https://www.3ieimpact.org/sites/default/files/2019-05/SR24-education-review_2.pdf

World Bank. (2018). *World development report 2018: Learning to realize education's promise*. https://www.worldbank.org/en/publication/wdr2018

World Bank. (2020). *World bank country and lending groups*. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Authors

DAVID K. EVANS is a senior fellow at the Center for Global Development. His research areas include education in low- and middle-income countries, teacher policy, and girls' education.

FEI YUAN is a doctoral candidate at the Harvard Graduate School of Education who researches education policy in low- and middle-income countries, program evaluation, systematic review, and girls' education.

Manuscript received February 18, 2021
Revision received October 13, 2021
Accepted January 23, 2022

Supplemental Appendix for
"How Big Are Effect Sizes in International Education Studies?"

by David K. Evans and Fei Yuan

This appendix includes the following:

- *Appendix A: Sample construction.* This includes a description of the search strategy, inclusion and exclusion criteria, and the process of standardizing effects.
- *Appendix B: Additional analysis.* This includes additional tables and figures to complement the analysis in the main brief.
- *References.* This lists references for sources cited in this appendix.

Appendix A: Sample Construction

To understand the distribution of effect sizes in international education studies, we built a sample of 280 eligible studies. Appendix Figure A1 summarizes the process.

Literature search

The starting point for this review is the set of all underlying studies in 11 recent systematic reviews of evidence on how to improve educational outcomes in low-and middle-income countries (Kremer et al., 2013; Krishnaratne & White, 2013; Glewwe et al., 2013, Ganimian & Murnane, 2016; McEwan, 2015; Masino & Niño-Zarazúa,, 2016; Glewwe and Muralidharan, 2016; Asim et al., 2017; Snilstveit et al., 2015; Conn, 2017; J-PAL, 2017). Those reviews yielded a total of 499 studies that were conducted between 1980 and 2015.

To include studies that were available after 2015, we performed a complementary literature search between October 2017 and January 2018. (When we prepared this analysis in 2020, we updated data for 20 studies that had subsequently been published in journals.) Specifically, we searched Google Scholar, major development research institutions and conferences (Appendix Table A1), and a set of economics and education journals for applied research (Appendix Table A1). Specifically, we looked for papers or research reports posted between 2015 and 2017 containing the keywords "evidence" OR "education" OR "access" OR "learning" OR "enrollment" OR "dropout" OR "attendance" OR "score" OR "developing countries" OR "Africa" OR "Asia" OR "America" OR "South Asia" OR "Europe". The search yielded additional 19 studies. An anonymous referee recommended one additional study that met all of our inclusion criteria. In total, we examined 519 studies.

Inclusion criteria

In our review, we include studies that evaluate (1) direct education interventions (such as teacher professional development and providing learning materials), as well as two other classes of interventions that commonly report educational outcomes: (2) health interventions (such as providing deworming drugs and micronutrients) and (3) safety net interventions (such as cash transfers). Studies also need to report at least one access outcome (e.g., enrollment, dropout, attendance) or a learning outcome (e.g., literacy, numeracy, test score). We focus on programs that took place in pre-primary, primary and secondary schools. Non-academic skill development programs for adolescents or cognitive development programs in early childhood were not included.

In terms of research methods, we restrict our analytic sample to studies that evaluated the underlying program using a valid counterfactual. Particularly, we only include studies that use an experimental design (RCT) or a quasi-experimental design (fixed effects, difference-in-differences, regression discontinuity, or instrumental variable) in the evaluation of effects. Studies that relied solely on matching (including propensity score matching) were excluded from our main analysis due to a lack of exogenous variation in treatment, although we do provide analysis comparing matching results with other strategies (compare Appendix Table B1 and B4, both panels A and B).

After the initial review, 181 studies did not meet our inclusion criteria, which yielded a sample of 338 studies for an in-depth review.

Coding effect sizes

In the in-depth review, we extracted and coded the effect sizes of educational outcomes in each study. When multiple outcomes were reported for the same intervention or program, we included all of them. However, we only coded effect sizes for both boys and girls and excluded gender-specific or other heterogeneous outcomes. (Studies that only reported gender-specific outcomes were excluded.) For learning outcomes, whenever available, we coded subject-specific outcomes instead of composite test scores.

Not all studies reported effect sizes. For studies that only reported point estimates, we convert them into standardized effect sizes or Cohen's *d*, following Borenstein et al.(2009).

$$d = \frac{D}{S_{pooled}} = \frac{\overline{Y_T} - \overline{Y_C}}{S_{pooled}} \quad (1)$$

where D is the raw mean difference between a treatment group ($\overline{Y_T}$) and a control group ($\overline{Y_C}$) at the follow-up, and S_{pooled} is the pooled standard deviation for the treatment and control groups combined. When S_{pooled} was not directly reported, it was calculated using equation (2) (Borenstein et al., 2009)

$$S_{pooled} = \sqrt{\frac{n_T n_C}{n_T + n_C}} SE_D \quad (2)$$

where n_T and n_C are the sample sizes in the treatment and control groups at the follow-up, and SE_D is the reported standard error of D.

61 studies did not report sufficient data for us to calculate effect sizes and hence were excluded, bringing us to a sample of 277 studies. For the analysis of this paper, we use reading or math test scores as the primary learning outcomes, and enrollment, attendance, and dropout as the primary access outcomes. Thus, studies that do not report any of these outcomes were excluded, yielding a final analytical sample of 234 studies.

Figure A1: Flow diagram summarizing sample construction

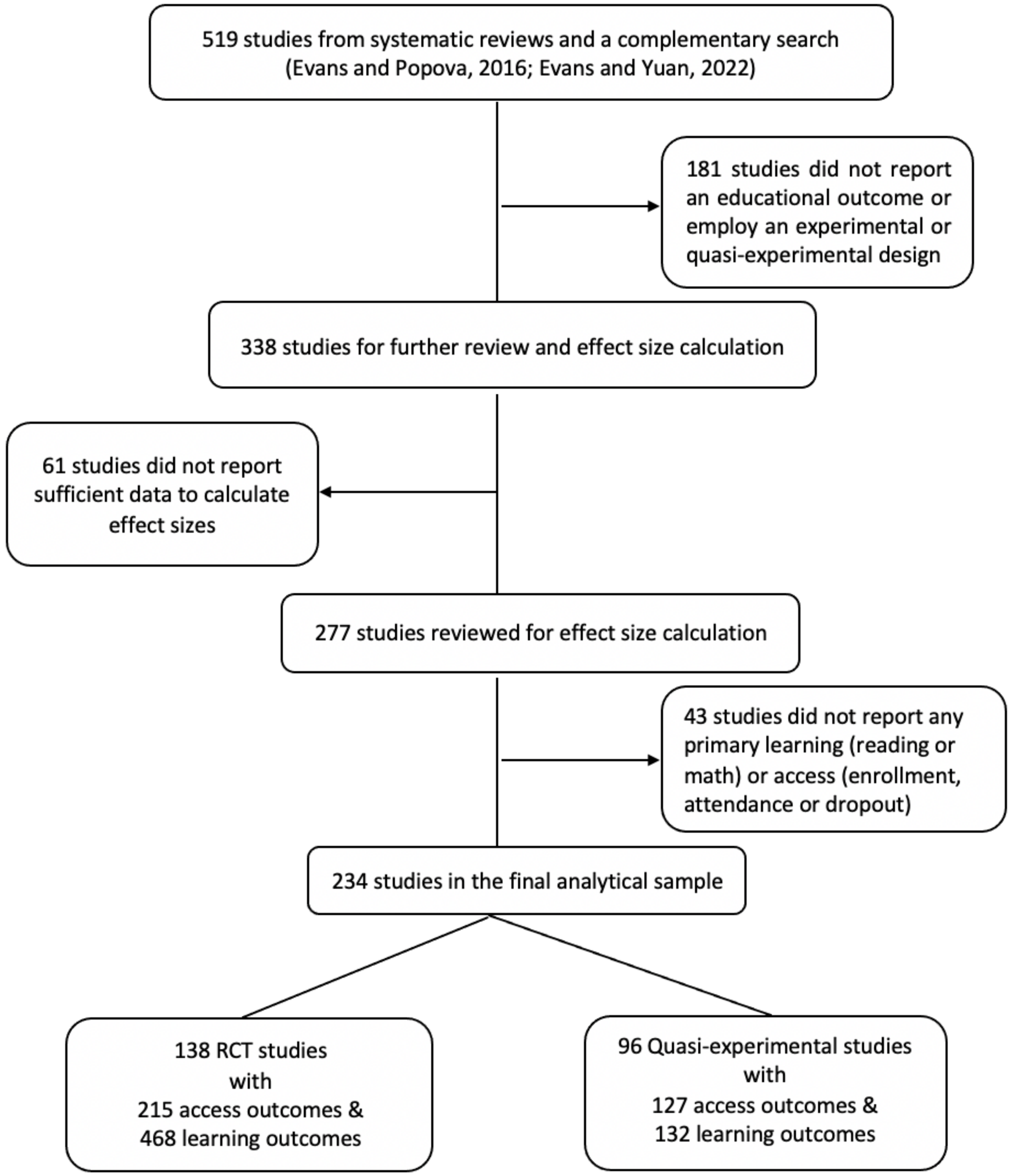


Table A1: Search locations

Locations	
Major development research institutions and conferences	• 3ie impact evaluation database, • RISE Programme Conference 2017 • CASE 2017 Conference • the Abdul Latif Jameel Poverty Action Lab at MIT • Innovations for Poverty Action • the Inter-American Development Bank • the National Bureau of Economic Research • RTI International • the Rural Education Action Program at Stanford University • the World Bank • the IZA Institute of Labor Economics
Economics and education journals	• American Economic Review • American Economic Journal: Applied Economics • American Economic Journal: Economic Policy • Comparative Education Review • Economic Development and Cultural Change • Journal of Development Economics • Journal of Development Effectiveness • International Journal of Educational Development • Journal of Human Resources • Quarterly Journal of Economics • The Economics of Education Review • World Bank Economic Review • Journal of Public Economics • Journal of Human Capital

Appendix B: Additional analysis

Table B1: Distribution of impacts across quasi-experimental studies

Panel A: Learning effect sizes

	Overall	Subject		Sample size			
		Reading	Math	1st quartile (<=1480)	2nd quartile (1480, 3225]	3rd quartile (3225,10363]	4th quartile (>10363)
Mean	0.15	0.18	0.11	0.23	0.18	0.06	0.12
Standard deviation	0.37	0.43	0.28	0.54	0.30	0.33	0.14
P1	-0.76	-0.78	-0.72	-0.78	-0.16	-0.72	-0.01
P10	-0.16	-0.20	-0.09	-0.43	-0.07	-0.27	0.00
P20	-0.07	-0.08	-0.02	-0.18	-0.01	-0.10	0.02
P30	0.00	-0.01	0.01	-0.08	0.01	-0.04	0.02
P40	0.02	0.01	0.02	0.03	0.02	-0.01	0.05
P50	0.05	0.05	0.05	0.11	0.05	0.01	0.06
P60	0.08	0.08	0.09	0.24	0.07	0.03	0.08
P70	0.19	0.22	0.16	0.47	0.19	0.07	0.14
P80	0.38	0.58	0.24	0.68	0.56	0.19	0.25
P90	0.72	0.83	0.40	1.15	0.77	0.48	0.40
P99	1.55	1.73	1.12	1.73	0.86	1.12	0.44
# of effect sizes	132	79	53	36	35	37	24
# of studies	43	36	37	10	10	12	12

Notes: Overall includes only reading and math effect sizes. The distribution of sample size is based on all quasi-experimental studies, not just those with reading or math outcomes; reported effect sizes by sample size quartile only include reading and math outcomes. Source: Authors' construction.

Panel B: Access effect sizes

	Overall	Subject			Sample size			
		Enrollment	Attendance	Dropout (Abs. value)	1st quartile (<=1480)	2nd quartile (1480, 3225]	3rd quartile (3225,10363]	4th quartile (>10363)
Mean	0.11	0.12	0.12	0.01	0.21	0.16	0.12	0.09
Standard deviation	0.22	0.25	0.16	0.08	0.48	0.29	0.35	0.13
P1	-0.20	-0.12	-0.05	-0.24	-0.78	-0.16	-0.72	-0.07
P10	-0.01	0.00	-0.01	-0.11	-0.40	-0.07	-0.15	0.00
P20	0.01	0.02	0.03	-0.02	-0.14	-0.01	-0.06	0.01
P30	0.03	0.03	0.05	0.00	-0.05	0.01	0.00	0.02
P40	0.05	0.05	0.06	0.01	0.09	0.02	0.01	0.02
P50	0.06	0.07	0.08	0.02	0.11	0.05	0.06	0.05
P60	0.08	0.08	0.09	0.02	0.26	0.06	0.12	0.07
P70	0.10	0.12	0.10	0.03	0.39	0.14	0.15	0.10
P80	0.15	0.16	0.23	0.05	0.57	0.32	0.18	0.14
P90	0.23	0.25	0.36	0.10	0.82	0.72	0.40	0.37
P99	1.65	1.98	0.68	0.16	1.73	0.86	1.59	0.45
# of effect sizes	127	86	25	16	51	40	60	47
# of studies	64	43	19	14	16	12	21	25

Notes: Overall includes enrollment, attendance and dropout (absolute value) effect sizes. The distribution of sample size is based on all quasi-experimental studies, not just those with enrollment, access or dropout outcomes; reported effect sizes by sample size quartile only include enrollment, access or dropout (absolute value) outcomes. Source: Authors’ construction.

Table B2: Distribution of learning effect sizes by assessment score type across RCT studies

	Composite test score	Standardized test score	Passed an exam	Cognition or social emotional learning
Mean	0.11	0.06	0.07	0.10
Standard deviation	0.14	0.22	0.08	0.12
P1	-0.29	-0.61	-0.09	-0.06
P10	-0.06	-0.24	-0.06	-0.06
P20	-0.02	-0.03	-0.01	-0.01
P30	0.02	-0.01	0.01	0.03
P40	0.07	0.03	0.05	0.06
P50	0.11	0.05	0.10	0.09
P60	0.16	0.12	0.11	0.13
P70	0.19	0.18	0.12	0.14
P80	0.21	0.28	0.14	0.17
P90	0.27	0.31	0.19	0.28
P99	0.57	0.34	0.20	0.45
# of effect sizes	74	16	13	18
# of studies	29	7	7	8

Notes: This table draws on 162 RCT studies that were eligible for effect size calculation (see Appendix A for details about eligibility). This is larger than the 96 RCTs with math and learning effect sizes because some studies only reported these alternative test results. Standardized test scores refer to standardized assessments at various levels, such as state, region or district. There are fewer than 10 effect sizes for science or social science in the sample, thus the distributions are not reported here. Source: Authors' construction.

Table B3: Distribution of learning effect sizes by assessment characteristics across RCT studies

	Mean	Standard deviation	P1	P10	P25	P50	P75	P90	P99	# of effect sizes	# of studies
Oral	0.23	0.31	-0.15	-0.02	0.04	0.14	0.35	0.57	0.93	217	33
Written	0.08	0.15	-0.38	-0.05	0.00	0.07	0.15	0.25	0.58	185	51
Not reported	0.12	0.19	-0.22	-0.07	0.02	0.08	0.20	0.37	0.75	66	17
Individual	0.21	0.24	-0.17	-0.03	0.04	0.14	0.35	0.56	0.91	221	35
Group	0.08	0.14	-0.38	-0.05	0.00	0.07	0.15	0.25	0.58	185	51
Not reported	0.18	0.42	-0.22	-0.06	0.02	0.09	0.22	0.42	3.03	62	16
Multiple-choice	0.52	0.91	-0.04	-0.04	0.08	0.24	0.47	2.81	3.03	10	5
Open-ended	0.23	0.25	-0.16	-0.02	0.03	0.14	0.37	0.59	0.91	180	27
Not reported	0.09	0.16	-0.29	-0.05	0.01	0.08	0.15	0.25	0.63	278	65
Specific domain	0.19	0.25	-0.27	-0.09	0.02	0.14	0.33	0.50	0.92	156	23
Not a specific domain	0.14	0.25	-0.22	-0.04	0.01	0.08	0.17	0.37	0.81	312	76

Notes: This table draws on the 96 RCTs with math and reading outcomes. Source: Authors' construction.

Table B4: Distribution of effect sizes by publication type and evaluation method
Panel A: Learning effect sizes

	RCT journal	RCT working paper	Quasi-experimental journal	Quasi-experimental working paper	All quasi-experimental plus matching studies
Mean	0.13	0.18	0.13	0.19	0.16
Standard deviation	0.25	0.25	0.38	0.36	0.52
P1	-0.19	-0.27	-0.78	-0.72	-2.59
P10	-0.04	-0.04	-0.20	-0.16	-0.15
P20	0.00	0.01	-0.07	-0.04	-0.06
P30	0.03	0.03	-0.01	0.01	0.00
P40	0.06	0.07	0.02	0.01	0.02
P50	0.09	0.11	0.05	0.06	0.06
P60	0.13	0.16	0.07	0.09	0.13
P70	0.15	0.28	0.16	0.36	0.23
P80	0.22	0.35	0.25	0.62	0.40
P90	0.35	0.51	0.57	0.82	0.73
P99	0.77	0.91	1.73	0.86	1.73
# of effect sizes	257	211	85	47	201
# of studies	58	38	30	13	65

Notes: Effect sizes only include reading and math outcomes. Source: Authors' construction.

Panel B: Access effect sizes

	RCT journal	RCT working paper	Quasi-experimental journal	Quasi-experimental working paper	All quasi-experimental with matching
Mean	0.10	0.10	0.11	0.11	0.15
Standard deviation	0.13	0.15	0.18	0.25	0.44
P1	-0.10	-0.10	-0.24	-0.11	-0.19
P10	-0.01	-0.04	0.00	-0.03	-0.02
P20	0.02	-0.01	0.01	0.01	0.01
P30	0.04	0.01	0.03	0.02	0.03
P40	0.05	0.02	0.05	0.03	0.05
P50	0.08	0.05	0.06	0.06	0.06
P60	0.09	0.08	0.08	0.08	0.08
P70	0.12	0.14	0.10	0.10	0.10
P80	0.16	0.17	0.15	0.16	0.17
P90	0.26	0.35	0.29	0.24	0.27
P99	0.71	0.59	0.81	1.98	3.44
# of effect sizes	151	64	56	71	139
# of studies	53	20	33	31	77

Notes: Effect sizes only include enrollment, attendance and dropout (absolute value) outcomes. Source: Authors' construction.

Table B5: Distribution of impacts by region
Panel A: Learning effect sizes

Region		**Mean**	**Standard deviation**	**P1**	**P10**	**P25**	**P50**	**P75**	**P90**	**P99**	**# of effect sizes**	**# of studies**
Sub-Saharan Africa	RCT	0.19	0.25	-0.18	-0.09	0.00	0.13	0.33	0.50	0.91	176	29
	Quasi-experimental	0.37	0.49	-0.78	-0.16	0.01	0.38	0.73	1.09	1.73	50	12
South Asia	RCT	0.15	0.22	-0.54	-0.01	0.02	0.07	0.28	0.55	0.70	107	19
	Quasi-experimental	-0.05	0.21	-0.72	-0.37	-0.10	0.00	0.06	0.20	0.22	22	5
Latin America and the Caribbean	RCT	0.14	0.33	-0.35	-0.03	0.03	0.09	0.16	0.31	2.99	101	24
	Quasi-experimental	0.04	0.15	-0.42	-0.08	-0.01	0.05	0.11	0.24	0.47	44	18
East Asia and Pacific	RCT	0.12	0.14	-0.11	-0.03	0.01	0.08	0.18	0.26	0.75	83	24
	Quasi-experimental	0.04	0.07	-0.10	-0.06	0.01	0.03	0.06	0.16	0.19	13	7
Europe and Central Asia	RCT	-	-	-	-	-	-	-	-	-	0	0
	Quasi-experimental	-0.19	0.02	-	-	-	-	-	-	-	3	1
Middle East and North Africa	RCT	-	-	-	-	-	-	-	-	-	1	1
	Quasi-experimental	-	-	-	-	-	-	-	-	-	0	0

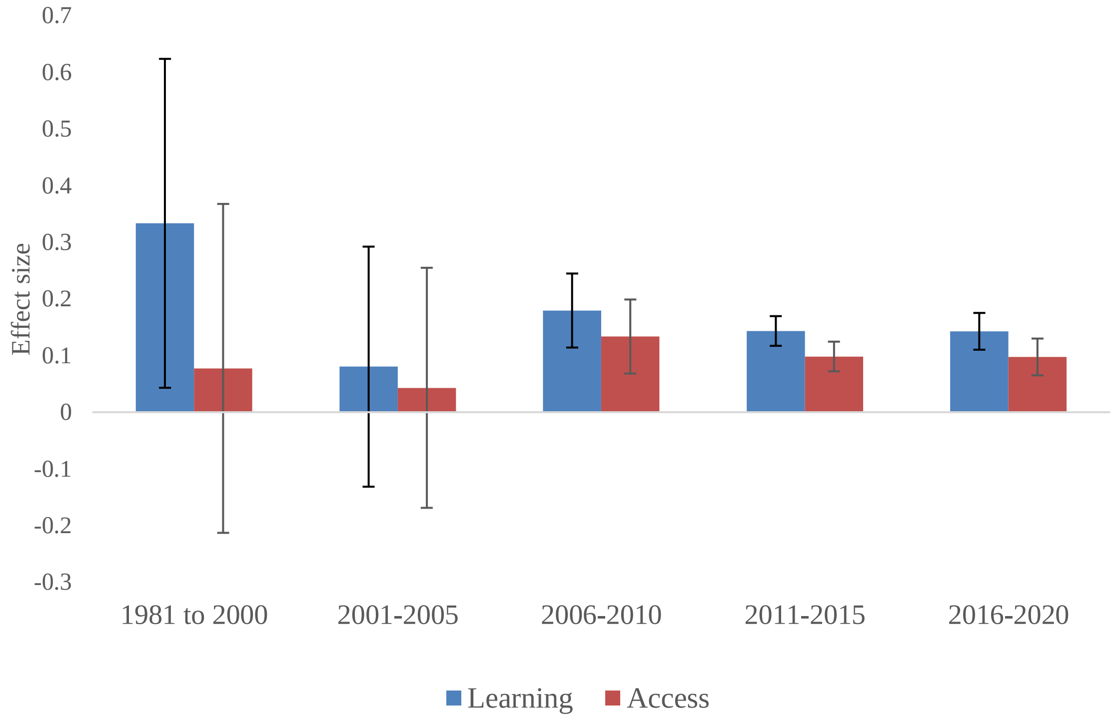
Notes: Region groups follows World Bank Country and Lending Groups (2020). Effect sizes only include reading and math outcomes. Source: Authors' construction.

Panel B: Access effect sizes

Region		Mean	Standard deviation	P1	P10	P25	P50	P75	P90	P99	# of effect sizes	# of studies
Sub-Saharan Africa	RCT	0.13	0.17	-0.10	-0.02	0.02	0.07	0.17	0.42	0.62	87	26
	Quasi-experimental	0.17	0.43	-0.05	-0.02	0.03	0.06	0.12	0.48	1.98	21	14
South Asia	RCT	0.08	0.09	-0.06	0.00	0.02	0.07	0.14	0.16	0.46	37	13
	Quasi-experimental	0.06	0.11	-0.24	-0.03	0.01	0.05	0.09	0.21	0.45	44	14
Latin America and the Caribbean	RCT	0.09	0.14	-0.09	-0.05	0.03	0.08	0.11	0.20	0.80	42	19
	Quasi-experimental	0.11	0.16	-0.05	0.00	0.02	0.07	0.13	0.27	0.81	50	29
East Asia and Pacific	RCT	0.07	0.11	-0.12	-0.07	0.00	0.06	0.13	0.22	0.37	41	14
	Quasi-experimental	0.19	0.20	0.01	0.02	0.05	0.13	0.23	0.63	0.68	11	6
Europe and Central Asia	RCT	-	-	-	-	-	-	-	-	-	-	-
	Quasi-experimental	-	-	-	-	-	-	-	-	-	1	1
Middle East and North Africa	RCT	0.14	0.11	0.01	0.01	0.05	0.12	0.26	0.30	0.30	8	1
	Quasi-experimental	-	-	-	-	-	-	-	-	-	-	-

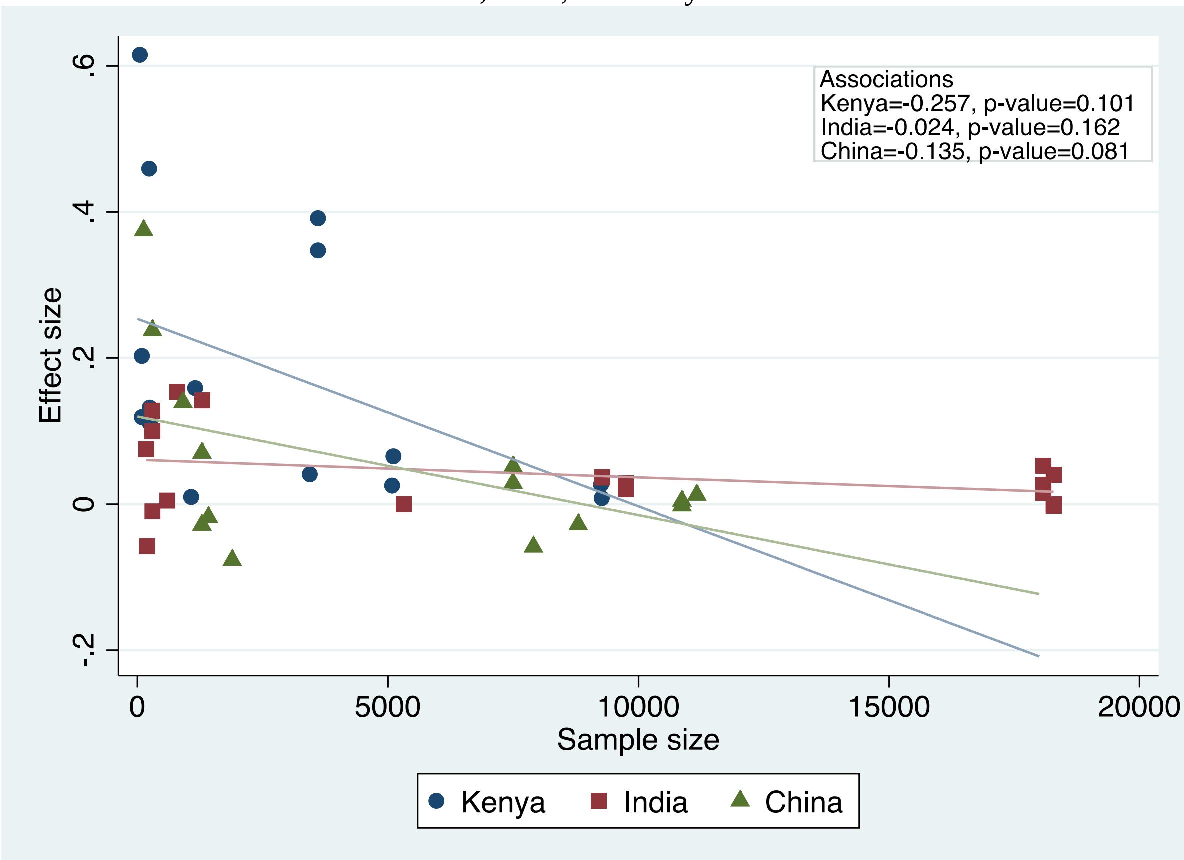
Notes: Region groups follows World Bank Country and Lending Groups (2020). Effect sizes only include enrollment, attendance and dropout (absolute value) outcomes. Source: Authors' construction.

Figure B1: Average effect sizes for learning and access outcomes over time



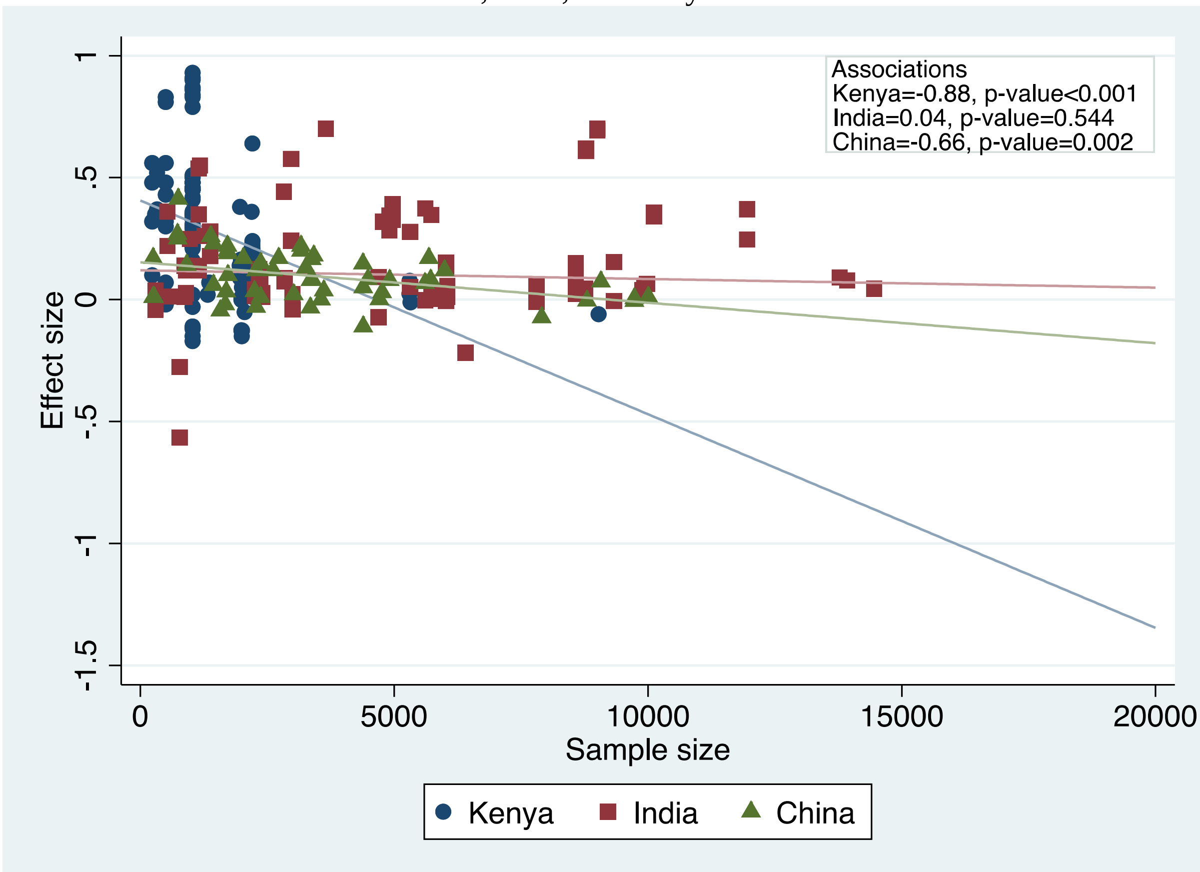
Notes: Learning effect sizes only reading and math outcomes. Access effect sizes only include enrollment, attendance and dropout (absolute value) outcomes. The literature search was completed in January 2018, but the database was later updated in 2020 to reflect the most recent publication version: 20 studies were published in a journal after 2018. Source: Authors' construction.

Figure B2: The association between sample size and effect size for access outcomes among randomized controlled trials in China, India, and Kenya



Notes: Effect sizes only include enrollment, attendance and dropout (absolute value) outcomes. Source: Authors' construction.

Figure B3: The association between sample size and effect size for learning outcomes among randomized controlled trials in China, India, and Kenya



Notes: Effect sizes only include reading and math outcomes. Source: Authors' construction.

References

Asim, S., Chase, R. S., Dar, A., & Schmillen, A. (2017). Improving learning outcomes in South Asia: Findings from a decade of impact evaluations. *The World Bank Research Observer, 32*(1), 75-106.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470743386

Conn, K. M. (2017). Identifying effective education interventions in sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research, 87*(5), 863-898.

Evans, D. K., & Popova, A. (2016). What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews. *The World Bank Research Observer*, *31*(2), 242-270.

Evans, D. K., & Yuan, F. (2022). What we learn about girls' education from interventions that do not focus on girls. *The World Bank Economic Review*, *36*(1), 244-267.

Ganimian, A. J., & Murnane, R. J. (2016). Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, *86*(3), 719-755.

Glewwe, P., Hanushek, E.A., Humpage, S.D., & Ravina, R. (2013). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. In *Education Policy in Developing Countries*. University of Chicago Press.

Glewwe, P., & Muralidharan, K. (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education* (Vol. 5, pp. 653-743). Elsevier.

J-PAL. (2017). *Roll Call: Getting Children into School*. Policy Bulletin.

https://www.povertyactionlab.org/sites/default/files/publication/roll-call-getting-children-into-school.pdf

Kremer, M., Brannen, C., & Glennerster, R. (2013). The Challenge of Education and Learning in the Developing World. *Science*, *340*(6130), 297–300. https://doi.org/10.1126/science.1235350

Krishnaratne, S., & White, H. (2013). *Quality education for all children? What works in education in developing countries*. International Initiative for Impact Evaluation (3ie).

Masino, S., & Niño-Zarazúa, M. (2016). What works to improve the quality of student learning in developing countries?. *International Journal of Educational Development*, *48*, 53-65.

McEwan, P. J. (2015). Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, 85(3), 353–394. https://doi.org/10.3102/0034654314553127

Snilstveit, B., Stevenson, J., Philips, D., Vojtkova, M., Gallagher, E., Schimdt, T., Jobse, H., Geelen, M., Pastorello, M.G., et al. (2015). *Interventions for improving learning outcomes and access to education in low- and middle-income countries: A systematic review*. 3ie Systematic Review 24. https://www.3ieimpact.org/sites/default/files/2019-05/SR24-education-review_2.pdf

World Bank. (2020). *World Bank Country and Lending Groups*. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups