RISE Working Paper 20/049
September 2020

Women's Education May Be Even Better Than We Thought: Estimating the Gains from Education When Schooling Ain't Learning

Michelle Kaffenberger, Lant Pritchett

Abstract

Women's schooling has long been regarded as one of the best investments in development. Using two different cross-nationally comparable data sets which both contain measures of schooling, assessments of literacy, and life outcomes for more than 50 countries we show the association of women's education (defined as schooling and the acquisition of literacy) with four life outcomes (fertility, child mortality, empowerment, and financial practices) is much larger than the standard estimates of the gains from schooling alone. First, estimates of the association of outcomes with schooling alone cannot distinguish between the association of outcomes with schooling that actually produces increased learning and schooling that does not. Second, typical estimates do not address attenuation bias from measurement error. Using the new data on literacy to partially address these deficiencies, we find that the associations of women's basic education (completing primary schooling and attaining literacy) with child mortality, fertility, women's empowerment and the associations of men's and women's basic education with positive financial practices are three to five times larger than standard estimates. For instance, our country aggregated OLS estimate of the association of women's empowerment with primary schooling versus no schooling is 0.15 of a standard deviation of the index, but the estimated association for women with primary schooling and literacy, using IV to correct for attenuation bias, is 0.68, 4.6 times bigger. Our findings raise two conceptual points, First, if the causal pathway through which schooling affects life outcomes is, even partially, through learning then estimates of the impact of schooling will underestimate the impact of education. Second, decisions about how to invest to improve life outcomes necessarily depend on estimates of the relative impacts and relative costs of schooling (e.g. grade completion) versus learning (e.g. literacy) on life outcomes. Our results do share the limitation of all previous observational results that the associations cannot be given causal interpretation and much more work will be needed to be able to make reliable claims about causal pathways.

Keywords: returns to schooling, education, literacy, child mortality, fertility, empowerment, non-pecuniary outcomes

Women's Education May Be Even Better Than We Thought: Estimating the Gains from Education When Schooling Ain't Learning

Michelle Kaffenberger
Blavatnik School of Government, University of Oxford

Lant Pritchett
Blavatnik School of Government, University of Oxford

This is one of a series of working papers from "RISE"—the large-scale education systems research programme supported by funding from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), the Australian Government's Department of Foreign Affairs and Trade (DFAT), and the Bill and Melinda Gates Foundation. The Programme is managed and implemented through a partnership between Oxford Policy Management and the Blavatnik School of Government at the University of Oxford.

Please cite this paper as:
Kaffenberger, M. and Pritchett, L. 2020. Women's Education May Be Even Better Than We Thought: Estimating the Gains from Education When Schooling Ain't Learning. RISE Working Paper Series. 2020/049. https://doi.org/10.35489/BSG-RISE-WP_2020/049

Contents

1 Introduction

"Schooling" and "education" are widely treated as synonyms[1]. Studies claim to examine the impact[2] of "education" on outcomes (e.g. wages, economic growth, women's empowerment, child health, political participation) but actually only examine the empirical relationships of these outcomes with measures of schooling completed. If schooling completed and education (schooling plus learning) were tightly associated within and across countries conflating the two terms might be benign. Unfortunately, often "schooling ain't learning" (Pritchett, 2013). Across the more than 50 countries with Demographic and Health Surveys (DHS) or Financial Inclusion Insights surveys (FII) data on schooling and literacy, only half of adults with primary schooling completed (and no higher) could read (Kaffenberger and Pritchett, 2017). Moreover, the extent to which schooling is a reliable indicator of having gained literacy varies widely across countries: in Nigeria, only 10% of women who completed six years of schooling could read a simple sentence, while in Rwanda more than 90% could (Pritchett and Sandefur, 2017). This generally weak and widely varying connection between schooling and learning implies the impact of *education*, a word that we argue necessarily implies the acquisition of useful competencies, on outcomes cannot be directly inferred from any estimate of the impact of schooling which relies exclusively on time served or grade completed.

We use DHS data from 54 countries (and a total of 128 survey rounds) and FII data from

[1]We would like to thank Justin Sandefur for close collaboration and discussions over the years that led to this paper, his inputs and outputs have been critical to improving this work to where it is, but he is not implicated in its remaining weaknesses. We would also like to thank participants in a RISE workshop where it was presented for the suggestions, Deon Filmer for support on an earlier version of this work, and Clare Leaver for helpful comments and feedback.

[2]A big caveat right up front: we use the word "impact" to refer to the empirical association (either partial or total derivative) estimated from a multivariate regression framework, not because we (naively) believe observational data produces unbiased/consistent estimates of a causal impact or LATE (local average treatment effect), but because other circumlocutions for "impact" are so awkward and unwieldy. If the reader (reviewer/referee) wants to mentally search and replace our use of "impact" with "partial (or total, depending on context) derivative of y with respect to x as estimated from linear (though this linearity is inessential) multivariate regression using observational data" nothing about our argument will be affected. We try and use notation and language (and periodically repeat caveats) that make clear our use of the word "impact" is short-hand for a particular coefficient, from a particular model, estimated in a particular way, not an assertion of identification of the "true" causal impact.

10 countries to estimate the empirical associations between schooling (years completed) and a measure of learning (ability to read) with non-pecuniary adult outcomes. With DHS data, we investigate the associations of women's schooling and learning with child mortality, fertility, and an index of women's empowerment. With the FII data, we investigate the associations of men's and women's schooling and learning with an index of financial behaviors. Using so many countries allows us to explore both the "typical" finding across countries as well as the heterogeneity across countries.

Using these two separate data sources, with two different literacy tests, administered to two different subsets of national populations (women of child bearing ages only in the DHS versus all adults for the FII), across four life outcome variables across many countries produces five, remarkably consistent, empirical results.

First, the typical approach of estimating empirical associations based on observational data, using OLS,[3] with schooling alone (and no measure for learning), underestimates the association of *basic education* (defined as primary schooling plus basic literacy) on outcomes by a factor of three to four. For example, using typical OLS approaches which only include schooling as a proxy for education suggests achieving primary schooling (six years) is associated with a 9.7% reduction in fertility. Our preferred method, including both schooling and literacy and correcting for measurement error using instrumental variables, yields an estimated fertility reduction of basic education (primary school completion plus basic literacy) of 36% – more than three times greater.

Second, in our preferred IV estimates achieving literacy, conditional on schooling, has as large, or larger, association with outcomes as completing primary schooling conditional on literacy. For women's empowerment the association with achieving basic literacy is *four*

[3] The use of OLS for estimating empirical associations is not the "state of the art" but the "state of play" or "industry standard." There are very few attempts to use randomized control trial or other methods of causal identification (like using the onset of "free primary education" as a instrument) to establish the causal impacts of schooling on non-pecuniary outcomes (Duflo et al., 2019). The "industry standard" of hundreds, if not thousands, of studies, and hence the evidence on which the existing claims about the non-wage impacts of women's schooling are currently based, uses observational data and cross-tabulations or OLS (and with discrete variables of interest one can think of multi-variate OLS regression as just a modestly extended cross-tabulations) comparing outcomes.

times larger than that of completing primary schooling (conditional on learning).

Third, because the observed acquisition of literacy per year of schooling differs (massively) across countries the total impact of schooling (the impact of schooling itself plus the impact it has through the pathway of the learning it produces at the observed pace of learning) must differ as well if learning is any part of the causal pathway for the impact of schooling. Any estimate of the total LATE (Local Average Treatment Effect) of an additional year of schooling on outcomes–whether using techniques to identify an unbiased estimate of causal impact or just based on observational data–can be decomposed into a direct impact on outcomes of schooling itself (conditional on learning) and an impact on outcomes of schooling through the learning it produces. We show that the percent of women with the same level of years of schooling completed who are literate differs across countries by an *order of magnitude*. Even if the impact on life outcomes of schooling (conditional on learning) and the impact on life outcomes of learning (conditional on schooling) were constant across countries or contexts the differences in learning produced from schooling would still cause massive differences in the LATE on life outcomes of schooling. This point aligns with earlier work by (Oye et al., 2016) which uses the DHS data for fertility and child survival and finds that in countries where learning is higher, the gains in outcomes associated with school are also higher.[4]

Fourth, estimation method matters, a lot. The use of instrumental variables estimation techniques to correct for measurement error (*not* causal identification) leads to much larger estimates, and for three of the four outcomes the difference is much larger for literacy, consistent with a simple literacy assessment as a noisy measure of learning. Using IV versus OLS nearly doubles the estimated partial association of schooling with child mortality, but for literacy it increases it three-fold (and for fertility five-fold). This suggests any attempt to decompose the "impact of education" into the relative partial impacts of schooling and learning must grapple with measurement error, as in many cases learning estimates will be more attenuated (biased towards zero) by measurement error than schooling estimates.

[4]For example they find the gains in terms of child survival are about two-thirds larger in countries with the highest learning compared to those with the lowest.

We will be the first to detail the many limitations of the instrumental variable approach we adopt, but the large, and differential, differences between OLS and IV estimates illustrate the importance of grappling seriously with measurement error as it has critical implications for estimating the relative impacts of schooling versus learning, as the two are highly correlated.

Fifth, our estimates are descriptive, based on observational data, and are not causal. Given the relatively low explanatory power of our estimates of life outcomes and the very high correlations of schooling and learning robustness tests suggest small deviations of assumptions about, say, selectivity or reverse causation (e.g. that women who are more likely to achieve literacy are also more likely to achieve better life outcomes because of non-observed individual specific variables) could reverse our empirical results. However, rather than negating the conceptual points this suggests the need for further empirical research investigating the differentiated impact of schooling and learning on life outcome, as even RCTs that demonstrate casual impacts of extending schooling (e.g. Duflo et al. (2019)) cannot disentangle causal pathways between schooling and learning nor (as pointed out above) be extrapolated to other contexts.

These empirical findings are all relevant to policy decisions, as optimal allocation of effort (or funding) to expand years of schooling versus to improve learning per year necessarily depend on the relative costs and the relative life outcome benefits. While there is increasing evidence about the cost effectiveness of various "interventions" in increasing either schooling or learning per year of schooling, they are insufficient for informing policy without consideration of the impacts on life outcomes of each, as either simplistic assumption that all the benefits are accomplished just by time served in schooling, or that all of the benefits are completely captured by learning metrics, are likely to be false.

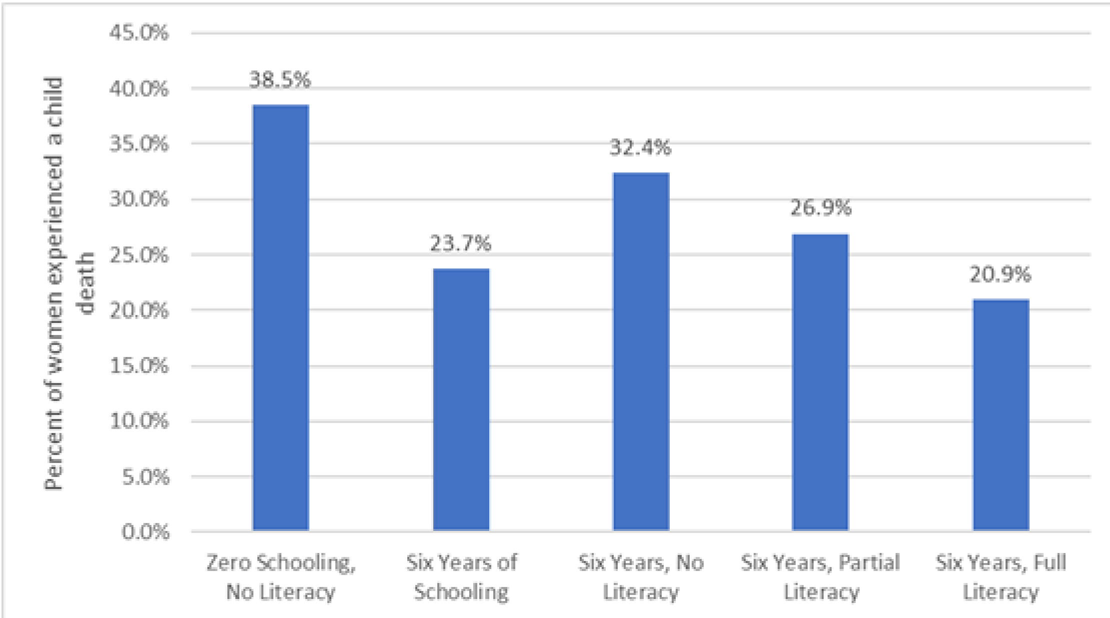
2 Why conceptual clarity on the impact of Schooling and the impact of education is important

2.1 The shortcomings of “schooling” as a proxy for “education”

The widespread availability of household survey data sets (many national and many with at least some cross-national comparability (e.g. DHS, MICS, LSMS, World Values Surveys, Young Lives) with measures of schooling completed and non-pecuniary outcomes (fertility, child mortality, political participation, attitudes, values, health care usage, child nutrition, child school attendance, etc.) has led to thousands of studies comparing life outcomes by individuals’ levels of completed schooling. This literature almost exclusively uses “schooling” as a synonym for “education.” Oft-cited estimates used to demonstrate the value of girl’s “education”, for example that child mortality declines 7-9% per year of women’s schooling (e.g. Cleland and Van Ginneken, 1988; Cochrane, 1980; Nations, 1985), and an analysis using 915 data sources from 219 countries which claimed that female “education” prevented 4.2 million child (under 5) deaths between 1970 and 2000 (Gakidou et al., 2010), are based only on measures of schooling. Similarly, observational studies have linked years of schooling to reductions in fertility via various pathways such as family size preference, age at first marriage, and contraceptive use (Martin, 1995), and to lower child malnutrition (Keats et al., 2017). Two recent systematic reviews examining causal links between female “education” and maternal and child health (Mensch et al., 2019) and sexual and reproductive practices (Psaki et al., 2019) included only one study (total, across both reviews) that included any measure of learning (both reviews acknowledge this shortcoming of the current literature).

A simple graph begins the illustration of the importance of not conflating schooling and education. Figure 1 is the cross-tabulation of child mortality (whether, among women who have ever had a child, a woman has ever experienced the death of a child) by the woman’s level of schooling and by the DHS measure of literacy using the data from 54 countries. The DHS literacy assessment classified women by whether they could read a full sentence

Figure 1: Fraction of women aged 15-49 who have experienced the death of a child, by schooling and literacy levels



Source: Authors' analysis of DHS micro-data, including $N = 854,766$ women who have ever given birth, from 54 countries.

without help, only read parts of a sentence, or not read the target sentence at all (more detail below). Of women with no schooling and no literacy (unschooled and uneducated) 38.5% have experienced a child death. Among women with six years of schooling complete but who could not read the sentence at all (schooled, but not educated) 32.4% had experienced a child death, only 6.1 percentage points lower than women without schooling or literacy. Among women with six years of schooling complete and who could read a sentence without help, that is, those with a basic education defined as primary schooling plus basic literacy, only 20.9% had experienced a child death, 17.6% percentage points lower than women with no formal schooling or literacy. The difference in child mortality between women with six years of schooling complete with and without reading is almost twice as big (32.4%-20.9%=11.5%) as the gap between women with no schooling who cannot read and those with six years of schooling and cannot read (38.5%-32.4%=6.1%).

The existing literature has demonstrated the differences in outcomes with and without schooling, which does not differentiate the impact by whether or not the schooling actually produced and learning. Hence estimates of the impact of schooling are a weighted average of the gains to schooling of women who achieved very different levels of learning from their schooling and hence are not estimates of the impact of women's education. As we see in Figure 1 the outcomes for women with basic education (schooling plus literacy), with only 20.9% experiencing a child death, are better than the outcomes for women with primary schooling that ignore measures of learning and pool together all women with primary schooling complete, 23.7%.

We extend this simple insight from the cross-tabulations to other indicators and more sophisticated estimation techniques to show that women's (and men's) education is even better than evidence for schooling alone suggests. Suppose that a life outcome (Y) for a specific woman (i) living in country c and locality j is a linear function of: her years of schooling completed (S), her extent of learning (L), and other factors about the woman that are in the data (Z, e.g. her age, whether she lives in a urban or rural location, a wealth index) plus everything else that affects outcomes besides S, L, Z:

$$Y^{i,c,j} = \alpha^c + \beta^c_{S|L,Z} * S_{i,j} + \beta^c_{L|S,Z} * L_{i,j} + \theta^c_{Z|L,S} * Z_{i,j} + \text{everything else} \tag{1}$$

We further assume that the learning achieved by a woman is linked to her schooling by a simple linear equation 2:

$$L^{i,c,j} = \eta^c + \gamma^c * S_{i,j} + \epsilon \tag{2}$$

where γ^c is the learning produced by a year of schooling in country c.

With this notation we can be clear about three different concepts, each of which could be called the "impact of schooling."

First, the partial derivative of the outcome with respect to schooling, which holds the extent of learning and the Zs fixed, is what we call the "partial" or "direct" impact of schooling

- the impact of schooling itself, holding all else including learning constant. In this notation this is $\beta^c_{S|L,Z}$, from equation 1.

Second, the *total* impact of schooling is the direct impact of schooling itself plus the impact schooling has on outcomes through the pathway of increasing learning. This is the total derivative of outcomes with respect to schooling after plugging equation 2 into equation 1, and is the sum of the partial impact ($\beta^c_{S|L,Z}$) and the pathway whereby schooling raises learning (γ^c) and hence, through the impact of learning on outcomes ($\beta^c_{L|S,Z}$), schooling leads to gains on outcome Y. This "total" impact of schooling is represented in equation 3.[5]

$$\text{Total impact of schooling} = \beta^c_{S|L,Z} + \gamma^c * \beta^c_{L|S,Z} * \tag{3}$$

If primary schooling consists of ΔS years of schooling at the level of learning produced by ΔS (γ^c) then the total impact of primary schooling on outcome Y is:

$$\text{Total impact of primary schooling } (Z \text{ fixed}) : \Delta Y = \beta^c_{S|L,Z} * \Delta S + \beta^c_{L|S,Z} * (\gamma^c) * \Delta S \tag{4}$$

The third possible understanding of the "impact of schooling" is if one defines the level of learning intended to be achieved during primary school and regarded as "basic" as ΔL. This is what we call the "impact of education" as it includes in the definition a given level of learning. The impact of a basic education, which is primary schooling plus a defined level of basic learning, is:

$$\text{Impact of basic education} : \Delta Y = \beta^c_{S|L,Z} * \Delta S + \beta^c_{L|S,Z} * \Delta L \tag{5}$$

The difference between the total impact of schooling and the impact of basic education

[5] This "total" ignores the effect of schooling and learning on the other factors, Z, for instance, if higher schooling raised incomes/wealth, and so this is really a "partial total" but we ignore those other pathways for simpler exposition.

is:

$$\text{Difference between impact of schooling and impact of education} = \beta^c_{L|S,Z} * (\gamma^c * \Delta S - \Delta L) \tag{6}$$

This difference, for any country c, is the gap between the learning *actually* produced by the S years of primary schooling ($\gamma^c * \Delta S$) and the level of learning regarded as "basic" and *intended* to be produced by the S years of primary schooling (ΔL), multiplied by the impact of learning on outcomes ($\beta^c_{L|S,Z}$).

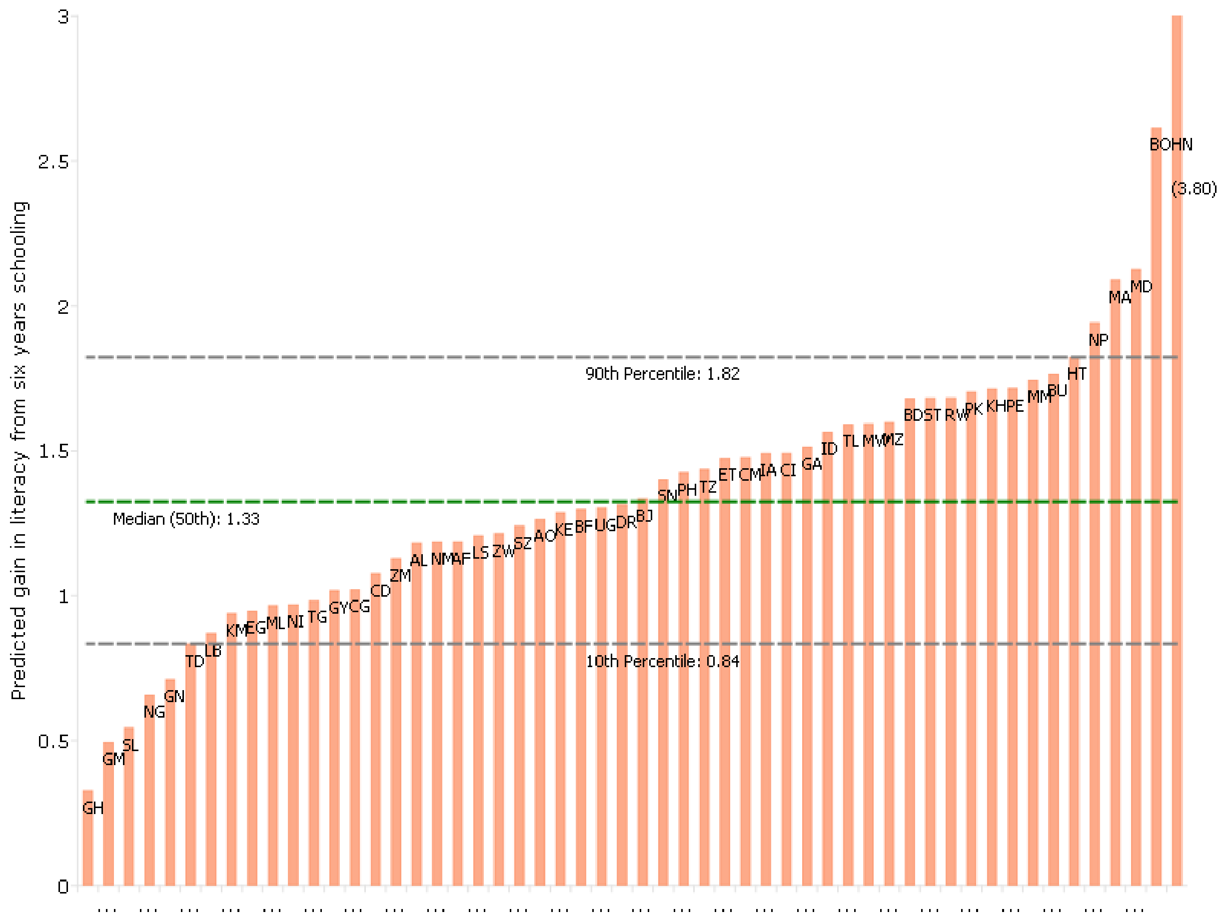
This simple equation has two substantive implications. One, the lower the learning gain from schooling (the smaller γ^c) the more the "total impact of schooling" underestimates the impact of education (for a given impact of learning on outcomes). Two, for a given learning gain (γ^c) the difference in the "total impact of primary schooling" and impact of basic education is larger, the larger the impact of learning on the outcome for a given level of schooling ($\beta_{L|S,Z}$).

If (i) schooling produced a (roughly) constant amount of learning (γ^c) across countries and (ii) that (common) degree of learning was (roughly) consistent with what was intended to be achieved in those years of schooling ($\gamma^c * \Delta S \approx \Delta L$), then, based on equation 6, the total impact of primary schooling and the impact of basic education would be conceptually distinct but empirically similar. However, neither of those premises are remotely true.

Drawing on previous research which produced descriptive learning profiles across countries from the FII and DHS data (Kaffenberger and Pritchett, 2017; Pritchett and Sandefur, 2017), we produce a rough and ready estimate of $\gamma^{c,R}$, where the super script R refers to reading ability.[6] We regress the categorical DHS literacy indicator (0-can't read at all, 1-woman could read the sentence with some help, and 2-she could read the simple sentence)

[6] Here and in subsequent sections we shift from an index "L" for "learning" to an index "R" when we are using our specific measures of "reading" to emphasize that estimated coefficients are dependent on the exact measures used. $\beta_{R|S,Z}$, the coefficient for just reading, may be smaller than $\beta_{L|S,Z}$ where "L" is a broader measure of learning. Hence, conversely, the "direct" impact of schooling would be larger (e.g. $\beta_{S|R,Z} >> \beta_{S|L,Z}$) as $\beta_{S|R,Z}$ is not conditional on the other measures of learning.

Figure 2: The predicted gain in literacy from six years of schooling varies by an order of magnitude across countries



Source: Authors' analysis of DHS micro-data.

on years of schooling. Figure 2 shows the predicted literacy for a woman with six years of schooling (but no higher) for each country.[7] Estimates of $\gamma^{c,R}$ range from a 90th percentile of 1.82 (most women with six years complete can read the sentence without help (R=2)) while the 10th percentile is only .84, implying many women with six years of schooling either cannot read at all (R=0) or need help (R=1).

The combination of Equation 6 and Figure 2 has two implications.

First, if the impact of schooling on outcomes is substantially through learning then, since γ^c differs across countries, the total impact of schooling (Equation 4) across countries will differ massively, even for countries for which the impact of education (Equation 5) is identical. For instance, if two countries had the same partial impact of schooling ($\beta_{S|L,Z}$) on outcomes, and the same impact of learning ($\beta_{L|S,Z}$) on outcomes, they would, by definition, have the same impact of basic education. However, to the extent that schooling produces different amounts of learning per year in these two countries (γ^c differs) their total impact of schooling would differ. For example, suppose that all of the impact on outcomes of schooling was through learning and countries A and B had the same impact on outcomes of learning, then the impact of schooling in country A versus country B would just be the ratio γ^c_A/γ^c_B. Using the results in Figure 2 this would imply the 90th percentile learning country would have an estimated impact of schooling of 2.16 (=1.82/.84) times higher than the 10th percentile learning country. There cannot be "external validity" across contexts/countries in estimates of the "total impact of schooling" (no matter how those are estimated) on *any* life outcome if learning per year varies, and it does[8].

[7]The graph shows the average coefficient for each country across all its available survey rounds. The coefficients, $\gamma^{c,R}$, are reassuringly stable across the survey rounds for each country. A simple decomposition of variance across the 115 survey rounds for the 40 countries with more than one survey round shows that 95 percent of the total variation in the estimates of $\gamma^{c,R}$ is due to country specific effects and only 5 percent due to survey round variations for the same country.

[8]Even well identified causal estimates of the impact of school cannot overcome these as identifying the channels through which schooling impacts outcomes requires estimates of γ^c and $\beta_{L|S,Z}$. For instance, Breierova and Duflo (2004) exploit variation from a nationwide school construction program in Indonesia to recover causal estimates of the impact of increased parental schooling on child mortality. Similarly, introduction of Universal Primary Education in Nigeria in 1976 and Uganda in 1997 provided researchers with a source of exogenous change; based on this analysis Osili and Long (2008) suggests that increasing female schooling by one year reduces early fertility in Nigeria, and Keats (2018) finds that women in Uganda

Second, estimates of the total impact of schooling will underestimate the impact of basic education to the extent that $\Delta L > \gamma^c \Delta S$. If we take the standard of basic education to be literacy (R=2) for all then, given equation 6, the impact of schooling will be lower than the impact of education in the median country by:

$$\beta^c_{L|S,Z} * (1.33 - 2) = -.67 * \beta^c_{L|S,Z} \tag{7}$$

For further discussion of the implications of this form of omitted variables bias, see Appendix A.

2.2 Implications for investing in education

This conceptual clarity matters because decisions about how to invest in education to improve pecuniary (wages, incomes) or non-pecuniary (child mortality, empowerment) life outcomes necessarily depend on an understanding of the channels whereby schooling has its impact on outcomes. Without estimates of the relative impacts of schooling (grade completion) versus learning (based on some measure) on life outcomes the estimates commonly produced of the impact of and cost effectiveness of various interventions in either raising schooling completed (at existing learning per year) or raising learning are, at best, incomplete guidance to optimal investments.

Suppose there was an intervention (program, project, policy) that at marginal cost c^S could raise girl's schooling by one year and another intervention that at marginal cost c^γ

with more schooling prefer to have fewer children, delay having their first child, and reduce overall fertility at any age, while investing more in their children's health. Similarly recent work in Ghana of a program to extend girl's enrollment in secondary school estimates the impact on fertility and other outcomes Duflo et al. (2019). But, no matter how well identified or precise, the estimates of the total impact of school, even cleanly identified causal estimates, do not provide any information on the impact of *learning* ($\beta_{L|S,Z}$) and hence of education - or what could be achieved if schooling produced learning. Moreover, even if a randomized or experimental design produces clean identification of the impact of schooling, and even if the study includes data on learning (e.g. literacy) this still would not provide well-identified estimates of the causal pathways (γ^c.) Further, such estimates of the causal impact of schooling on mortality in Indonesia depends on $\gamma^{Indonesia}$ and hence, even if $\beta_{S|L,Z}$ and $\beta_{L|S,Z}$ themselves were constant across all countries the estimate from Indonesia cannot be used to estimate the impact of schooling in countries with much lower (e.g. in the DHS data for instance, Ghana, Nigeria) or much higher learning countries unless $\beta_{L|S,Z}$ is zero.

could raise the learning per year of schooling, γ. We want to know which is the most cost effective for increasing the impact of education on a life outcome (child mortality, wages, empowerment, etc.). The standard optimizing decision rule is equate the marginal benefit per dollar across the two possible interventions, where we set the cost of an additional year of schooling (c^S) to equal 1 as a normalization.

$$\left(\frac{MB}{MC}\right)^{\gamma} = \left(\frac{\beta_{L|S,Z} * S}{C^{\gamma}}\right) \tag{8}$$

$$\left(\frac{MB}{MC}\right)^{S} = \left(\frac{\beta_{S|L,Z} + \beta_{L|S,Z} * \gamma}{C^{S}(\equiv 1)}\right) \tag{9}$$

Equations 8 and 9 imply that the cost of a learning-increasing intervention (relative to the cost of an incremental year of schooling) that would equalize the MB per dollar of the two interventions in producing a particular outcome Y is:

$$MC^{*}(\gamma) = \frac{\beta_{L|S,Z} * S}{\beta_{S|L,Z} + \beta_{L|S,Z} * \gamma} \tag{10}$$

Equation 10 is in terms of scaled quantities (the β are in units specific to the particular outcome) and so cannot be interpreted directly, but the equation has intuitive features. The higher the level of schooling, S, the larger the gains from increasing learning per year (γ) as this learning happens for more years and hence the higher the marginal cost of a learning increasing intervention could be and still be optimal. If none of the causal impact of schooling is through L (the measure of learning) then $\beta_{L|S,Z}$ is zero and the optimized cost of a learning-increasing intervention would have to be zero. Conversely, if none of the causal impact is the partial effect of schooling, conditional on the (measure of) learning, ($\beta_{S|L,Z} = 0$), then all the impact is through learning and Equation 10 reduces to $\frac{S}{\gamma}$ so that the higher S or the lower γ the higher the marginal cost of an optimal learning intervention.

There is a massive, and rapidly expanding, literature creating estimates of the impact and cost-effectiveness of various interventions using rigorous methods for estimating causal impacts of interventions on schooling or on learning per year of schooling.[9] This literature is useful, but without measurements of the relative impacts of schooling and learning on outcomes, both pecuniary and non-pecuniary, one cannot use such impact and cost-effectiveness estimates to make recommendations across the two classes of interventions. For instance, Barrera-Osorio et al. (2018) evaluated two different scholarships given to fourth grade students in Cambodia, one merit based and one needs based, which were awarded in 2008. In their long-term follow-up, nine years after the scholarship began, they found that while both programs had roughly equal effects on additional schooling, only the merit-based scholarship had any impact on learning or on any other measured life outcome. An evaluation of these alternative scholarship designs solely on the basis of additional S would have regarded them as equally cost effective in units of S gained per dollar. But a fuller analysis tracing through to learning and to outcomes revealed one design ("merit") produced more S *and* more L and led to impact on outcomes whereas the other design ("need") produced only more S but not more L (hence less than would have been expected from the additional schooling) and had no demonstrable impact on life outcomes and hence was massively less cost-effective at producing improved outcomes.

In section 6 we return to this point and illustrate these conceptual points below using our empirical estimates of $\beta_{L|S,Z}$, $\beta_{L|S,Z}$, S, and γ^c.

3 Data on Schooling, Literacy, and Outcomes

Our estimates advance the literature by using household survey data that include: a measure of schooling, life outcome variables, individual and household level co-variates, and, most importantly, an enumerator-administered literacy test. This section describes each of those for our two primary data sources, DHS and FII.

[9]See, for example: Dhaliwal et al. (2013), Ganimian and Murnane (2016), Glewwe and Muralidharan (2016), though overall cost-effectiveness as part of impact evaluation is still not common Brown and Tanner (2019)

3.1 DHS and FII Data on Schooling and Literacy

The DHS and FII are nationally representative sample household surveys which use a common questionnaire and each produce comparable data across multiple developing countries.

We use the 128 DHS survey rounds from 54 different countries which contain the literacy assessment introduced around 2000. The DHS survey chooses one woman aged 15 to 49 (reproductive age) from each sampled household to complete a detailed women's questionnaire, which contains the literacy assessment. The DHS survey instrument asks each sampled woman whether she attended school and if so, the highest level she attended (primary, secondary, or tertiary), and also asks the highest grade she attended within the reported level. We use this self-reported highest grade attained as our measure of schooling.

The literacy assessment is administered only to women who report completing primary school or less as their highest level of schooling. Women taking the literacy assessment are asked to read a sentence from a card. The enumerators are provided with cards in the variety of languages they expect to encounter and each woman is allowed to choose the language she wishes to read. This is therefore not a test of literacy in the dominant national language but of a woman's ability to read in *any* language of her choosing[10]. The card contains one simple sentence like:

- *Parents love their children.*
- *Farming is hard work.*
- *The child is reading a book.*
- *Children work hard at school.*

Enumerators code whether the woman could: (i) read the full sentence, (ii) read parts of the sentence only, or (iii) not read at all. We consider women who could read the full sentence to be "literate," as reading one simple sentence is already a low bar for literacy and

[10]The data report those for whom an appropriate language card was not available and this was typically a small percent. These women, by not having a literacy result also do not figure in our results.

those who could read "part" of a sentence may have only been able to read a single word.

Our analysis of the connections between schooling, literacy, and life outcomes is necessarily restricted to the subset of women with primary school or less and this complicates considerations of how selectivity affects the estimates of impact, discussed below.

The FII surveys are nationally representative surveys in ten low- and lower-middle income countries (Bangladesh, Ghana, India, Indonesia, Kenya, Nigeria, Pakistan, Rwanda, Tanzania, and Uganda) and include as respondents both men and women. We use the most recent rounds, collected in 2015, for each country[11].

The FII surveys ask respondents their highest level of schooling by category and we use the five categories: "no formal education," "primary education not complete," "primary education complete," "some secondary," "secondary complete" in our regression analysis. We exclude those who started or completed tertiary, a very small, and highly selected, part of the sampled population[12].

At end of the FII questionnaire respondents are asked if they consent or not to photographs taken by the enumerator being used in research materials. The respondents are asked to read the three-sentence consent paragraph,[13] and the enumerator selects the category that corresponds with the respondents reading ability: (i) can read the informed consent form fluently without help; (ii) read well but had a little help; (iii) struggled and had a lot of help; or (iv) was unable to read/asked interviewer to read. We define an FII respondent as "literate" if they could read the text without help. The FII administers the literacy test to all respondents and does not have the same selection issues as the DHS.

[11] More information on the surveys can be found here: finclusion.org

[12] The main econometric concern is the combination of the possibility there is a non-linear relationship between measured schooling and outcomes and that tertiary education is highly selective and therefore using a linear regression might cause the estimates to be leveraged up by this part of the sample and hence the linear estimates would not actually be a good estimate of the incremental benefit of moving from, say, 6th to 8th grade. We could have kept these observations in the sample and then allowed for non-linearity (e.g. allowed for splines in the impact terms) and focused on our range of interest, but it is simpler to just drop these observations.

[13] The exact English text from the Kenyan survey instrument is: "We would like to take some photographs of you and your household. We will include some of the photographs in our reports. We might also publish some of them online on our website." This text was translated into the relevant local languages.

The literacy rates as measured by the DHS and FII for women with similar levels of education are, reassuringly, similar in levels and strongly correlated across countries (Appendix B).

The OECD defines literacy as "understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge potential" (OECD, 2009). UNESCO defines literacy as the "ability to read and write with understanding a simple statement related to one's daily life. It involves a continuum of reading and writing skills, and often includes basic arithmetic skills." The DHS and FII assessments of just reading a sentence or short passage are a very low bar for literacy by these definitions. As one comparison point, the city of Jakarta, Indonesia participated in the OECD PIAAC (Programme for the International Assessment of Adult Competencies) assessment of adult literacy. In the PIAAC assessment 57% of adults 25-65 with less than upper secondary complete were classified as "below level 1" (the bottom code). In sharp contrast, 77% of those with less than secondary school complete were classified as literate by the FII (in the top two categories) and 75% of those without secondary education as literate by the DHS. Many of those the DHS and FII classify as literate are in the bottom code of assessed functional literacy by PIACC.

The literacy variables in both the FII and DHS data are categorical (and FII reports only highest level of schooling completed, not years of schooling, and so schooling is also categorical). We use the literacy variables as both a dependent and independent variable in linear regressions which imposes both cardinality and linearity on the categorical variable. Our checks revealed treating literacy as categorical was a reasonable approximation (e.g. goodness of fit did not fall much, and the effect on outcomes of moving from "none" to "some" literacy was roughly the same as from "some" to "full") and linearity is much easier to report and use.

In reporting regression results we re-scale the DHS linear regression schooling coefficients by 6, so the magnitude compares no schooling versus six years complete, roughly equivalent to primary schooling completion. We rescale the DHS coefficient on literacy by 2, so the magnitude is no literacy versus read without help. The FII schooling coefficient is scaled by 2 to compare no schooling to primary completion and the literacy coefficient by three to represent moving from the bottom to top category in the four category literacy scale. This enables the comparison of the schooling and literacy coefficients across the DHS and FII.

3.2 Outcome variables

DHS outcomes We analyze three life outcome variables from the DHS: (i) fertility, the woman's self-reported total live births, (ii) child survival rate, the number of living children divided by total number of live births, and (iii) a measure of women's empowerment.

Our measure of women's empowerment is a standard empowerment index (e.g. Kishor and Subaiya, 2008) of the first principle component of the following questions:

QI) (positive indicators) Whether the woman has any say in the following decisions :

- Her own healthcare
- Making large household decisions
- Visiting family or relatives
- What to do with money her husband earns

QII) (negative indicators) Whether the woman believes a husband beating or hitting his wife is justified if the wife:

- Goes out without telling him
- Neglects the children
- Argues with him
- Refuses to have sex with him
- Burns the food

QIII) Whether the woman believes a wife may refuse sex with her husband if he "has other women."[14]

[14]The empowerment index is estimated separately for each survey round and is normalized within each survey round to mean of zero, standard deviation 1. Hence the coefficients are comparable across countries

Financial behaviors From the FII surveys, we construct a financial behaviors index as the life outcome of interest. The original objectives of the FII surveys were to measure the uptake and use of financial products and services among the adult population in each country in order to identify potential needs for additional financial services. The surveys thus include several questions on use of services such as bank accounts, mobile money, insurance, and savings instruments as well as questions on financial behaviors such as saving for emergencies, paying bills on time, and planning how to spend money. We construct a principle components index summarizing these financial behavior indicators. We use binary indicators for use of financial services, including bank account use, mobile money account use, and having at least one type of insurance, all of which are common financial inclusion indicators. We also include an ordinal savings variable with values representing not saving, saving with informal financial tools (e.g. saving at home), and saving with formal financial tools (e.g. with a bank or mobile money) to indicate sophistication of savings behaviors. We then include indicators for respondents' money management behaviors; a binary indicator was included for agreement with each of the following statements:

- "I spend less than I make each month"
- "I have an emergency fund to cover unplanned expenses"
- "I pay my bills on time"
- "My savings are larger than my debts"
- "I am highly satisfied with my present financial condition"

And finally, an categorical variable represented answers to the question, "how often do you make a plan for how to spend your income?" with answer options of "always or most of the time", "sometimes", "rarely", or "never". The financial behaviors index was standardized to have a mean of zero and a standard deviation of one for each of the 10 surveys.

in standard deviation units but these may represent different "absolute" amounts.

4 OLS estimates of the impact of schooling (partial and total) and the impact of education

In this section we report OLS regression results with the four life outcome variables as dependent variables, first with schooling plus other covariates only with no measure of learning (literacy) and then with schooling, covariates, and the measure of reading ability. This produces simple OLS empirical estimates of: the partial impact of schooling, the total impact of schooling, and the impact of education. In the next section we take this a step further and run these regressions using an approach that addresses attenuation bias from measurement error (IV). Because measurement error likely has differential implications for measures of schooling and learning, approaches that correct for measurement error substantially affect estimated results.

Presenting our results is complicated as we do not pool the data across survey rounds but rather in this section and the next we report the results of 1332 distinct regressions (using two different specifications (with and without literacy), two different estimation techniques (OLS and IV), four different outcome variables (child mortality, fertility, empowerment, financial practices), for 128 survey rounds for fertility and child mortality, 67 survey rounds for empowerment, and 10 countries for financial practices). We characterize the distribution of each of the sets of results with a central tendency of the distribution–what is the "typical" value from the set of countries and the standard error of estimating the central tendency. We also report on the dispersion across the countries/rounds.

4.1 Meta-analysis weighting

Our estimate of the central tendency for each set (life outcomes, specification, and estimation technique) is the standard random effects meta-analysis formula for the aggregation of the results of different studies, as each regression can be thought of as the result of a separate study. The random effects formula allows the "true" coefficients of schooling and literacy on outcomes to differ across countries.

$$\beta^K = \sum_{i=1}^{N} \frac{\beta_i^K}{var(\beta_i^K + \tau^2)} \Big/ \sum_{i}^{N} \frac{1}{var(\beta_i^K + \tau^2)} \tag{11}$$

Where β^K is the weighted sum of betas for either schooling, literacy, or education (which is a linear combination of schooling and literacy), β_i^K is the coefficient from survey round i, $var\left(\beta_i^K\right)$ is the estimate of the variance of β_i^K. The τ^2 term accounts for the variation between studies (survey/years) in the random effects model. Each estimated coefficient β_i^K is weighted by the inverse of its variance plus τ^2, hence more precise estimates are given more weight than less precise estimates (this is important for the IV results as some survey rounds have very high variance).

We also report the standard error of estimation of the central tendency. As we see, these are very small relative to the estimates, producing very high powered rejections of a null hypothesis that the "typical" effect is zero, which is the result of a relatively large number of distinct estimates each with a moderate standard error. However, it is important to be clear that this is not a measure of either (a) the standard error of estimation for the typical estimate nor (b) the dispersion of the distribution of the estimates: it can be the case both that the central tendency has a very small standard error but the standard deviation (or other dispersion measure) of the estimates themselves is quite large.

We prefer the aggregation using the random effects formula for meta-analysis, but, given how many studies we have, pretty much any other reasonable method of estimating the central tendency of the distribution will produce similar results. The Tables in Appendix C show the results for fertility, child survival and women's empowerment aggregated using either the median (unweighted) average across survey rounds and the median is nearly always identical[15].

[15] Although at times the simple average differs from the median or weighted average. This illustrates the benefits of using a weighted average that takes into account the precision of the estimates as otherwise a country with a very low precision estimate can be an extreme observation and have disproportionate influence

4.2 OLS Estimates

The first column under each outcome variable in Table 1 reports average scaled results across survey rounds with meta-analysis random effects weighting for regressions with schooling alone (with co-variates). These are scaled to reflect the impact of six years of schooling (or "primary complete" for FII) versus no schooling. These results are similar to the bulk of the existing literature on women's schooling, which uses OLS regressions of outcomes on years of schooling complete which include some control variables (Z) but without any measure of literacy.[16]

In this base case/existing literature specification, completing primary school is associated in the typical country with an average reduction in fertility of 0.33, a 10% reduction; a reduction in child mortality of 2.3 percentage points, a 22% reduction off of the baseline of 10.7%; a 0.146 standard deviation increase in the women's empowerment index and, a 0.28 standard deviation increase in the financial behaviors index.

Given that we are aggregating over estimates from many country/survey rounds that, in the case of the DHS, cumulatively use millions of observations, the standard errors of the (weighted) average are very small relative to the estimates. The standard "t-tests" (ratio of estimate to standard error) are between 9.24 and 23 and hence the p-levels of the hypotheses of zero association of primary school completion and women's life outcomes in the typical country are astronomically small. Appendix D shows p-values across all survey rounds for the three DHS outcomes.[17]

on the estimate of the central tendency. This is particularly true for the IV results, as some have very large standard errors.

[16]Much of the general "gray" literature by development organizations and advocacy groups is even simpler just showing cross-tabulations of outcomes across levels of schooling (as in Figure 1) without any controls.

[17]There is justifiable recent scepticism about results with the classical/standard hypothesis testing approach of using a "10/5/1 percent significance level" (probability of type I error) as this (over)emphasis on a fixed, relatively large, p-level has led to problems like "p-hacking" and a "crisis of reproduciblity" in the aggregation of small N, moderate p-level studies. Some fields, like particle physics, have adopted standards where 3 sigma results are "evidence" but announcing the "discovery" of a new particle requires 5 sigma evidence, which is a 1 in 3.5 million probability of the observed results being generated by chance (e.g. type I error of falsely rejecting the null of "no particle"). Our Table 1 OLS results are all at least 9 sigma. Appendix D shows the standard p-hacking graphs for the DHS results, which, not surprisingly, as we estimate and report exactly the same functional form over all available survey rounds, show no evidence of p-hacking.

There is also considerable heterogeneity across the estimates (necessarily so as we show below) and the reported the 20th and 80th percentile of the country/survey round estimates for each outcomes shows a wide range. Tables in Appendix C give a variety of descriptive statistics of the results across the DHS survey rounds, like the number of the estimates for each outcome that are of the expected sign and significant, which, for instance, is 92 of 128 for child survival, 100 of 128 for fertility, and 44 of 67 for women's empowerment. Most of the remaining results are statistically insignificant and very few are both of an unexpected sign and statistically significant (1,2 and 1 for child survival, fertility and women's empowerment respectively).

Our "base case" results intended to replicate the standard empirical practice. Using four different indicators from two completely different data sources from many different countries we strongly confirm the established findings that in observational data a woman's level of schooling is strongly and robustly positively associated with a variety of non-pecuniary life outcomes.

As discussed in Section 2, these estimates fall short of estimating the association between basic education and outcomes, however. Estimates of the impact of schooling are not bad or biased estimates of the impact of education, they are not estimates of the impact of education at all.[18].

The second column of Table 1 for each outcome presents the meta-analysis aggregated OLS estimates and 20th and 80th percentile of the distribution of OLS estimates with reading included in the regressions.

The estimates for reading show that, even among women with the same schooling com-

[18]See Appendix A for discussion of interpreting these estimates in light of the omitted variables bias caused from omitting a measure of learning from the regressions. When the omitted variable, learning or reading, is strongly correlated with schooling the OLS estimates with schooling alone actually produce estimates very close to the "total impact of schooling". When combined with the results in the second column for each outcome variable that includes the measure of reading we have two estimates for the "total impact of primary schooling" for each survey round (country/year): the OLS estimate that has omitted variables bias (equation 14) and, using the estimates that include both schooling and literacy plus a bi-variate regression of literacy on schooling to estimate $\gamma^{c,R}$ (equation 4). The correlation between the two estimates (simple OLS and equation 4) is .998, .999 and .998 for child survival, fertility, and empowerment respectively.

Table 1: OLS regressions for women's life outcomes with schooling and schooling and reading

	Fertility (DHS)		Child survival (DHS)		Women's empowerment (DHS)		Financial Behaviors Index (FII)	
	OLS, Schooling only	OLS, Schooling and Reading	OLS, Schooling only	OLS, Schooling and Reading	OLS, Schooling only	OLS, Schooling and Reading	OLS, Schooling only	OLS, Schooling and Reading
Primary Schooling (mean, RE meta-analysis weights)	-0.329	-0.259	0.023	0.017	0.146	0.081	0.279	0.18
Std. Error of mean	0.018	0.02	0.001	0.001	0.013	0.014	0.03	0.034
20th percentile of results	-0.536	-0.473	0.013	0.004	0.059	-0.001	0.177	0.082
80th percentile of results	-0.146	-0.044	0.035	0.031	0.244	0.172	0.418	0.338
Reading (mean, RE meta-analysis weights)		-0.105		0.009		0.108		0.324
Std. Error of mean		0.015		0.001		0.013		0.017
20th percentile of results		-0.257		-0.001		0.023		0.254
80th percentile of results		0.05		0.022		0.214		0.382
Education, linear combination of primary schooling and reading (mean, RE meta-analysis weights)		-0.373		0.027		0.193		0.503
Std. Error of mean		0.018		0.001		0.015		0.037
20th percentile of results		-0.562		0.015		0.084		0.4
80th percentile of results		-0.194		0.046		0.315		0.582
Ratio estimate of mean of impact of education (primary schooling plus reading) to mean of primary schooling alone		1.13 (-.373/-.329)		1.19 (.027/.023)		1.32 (.193/.146)		1.8 (.503/.279)
Number of survey rounds	128	128	128	128	67	67	10	10

Note: All Regressions contain controls for age, age squared, age cubed, a wealth index, a rural/urban dummy, and dummies for DHS sampling regions. Schooling coefficients have been scaled to reflect primary school completion; reading coefficients have been scaled to reflect going from no reading to reading the sentence/passage without help.

pleted, a woman who can read (versus no reading) has average fertility lower by 0.11 births, child survival higher by 0.9 percentage points, an empowerment index is 0.108 standard deviations higher. Among men and women with the same schooling, those who can read have a financial behaviors index score higher by 0.32 standard deviations. The t-statistics for the hypothesis test that there is zero association of reading with life outcomes in the typical (meta-analysis weighted) country range from 7 to 18 and hence, as with schooling alone, reject the null of zero as the central tendency of the estimates at astronomical significance levels.

With estimates for schooling and reading we can calculate the impact of basic education. The basic education estimate in the second column for each outcome is the linear combination of the estimated impact of completing primary school plus the impact achieving reading (equation 5). A woman with basic education has average fertility lower by 0.373 births compared with a woman with no schooling and no literacy. Child survival is higher by 2.7 percentage points, and the index of women's empowerment is by .19 standard deviations. The financial behaviors index is higher by .50 standard deviations.

As we showed above must be the case when learning from primary school falls short of producing universal reading, the typical impact of basic education (IBE) is larger the typical impact of primary schooling (IPS): by 13 percent for fertility, 19 percent for child survival, 32 percent for women's empowerment and 80 percent for financial behaviors. Women's education is even better than we thought, because the conflation of women's schooling with women's education underestimates the benefits of education when schooling doesn't produce learning.

Using equation 6 for the difference between IBE and IPS, the estimates in Table 1, and the estimates of γ^c for each country from Section 2 we can calculate the gap for each country (or survey/round) as a ratio of the estimated IPS, which for the DHS outcomes is:

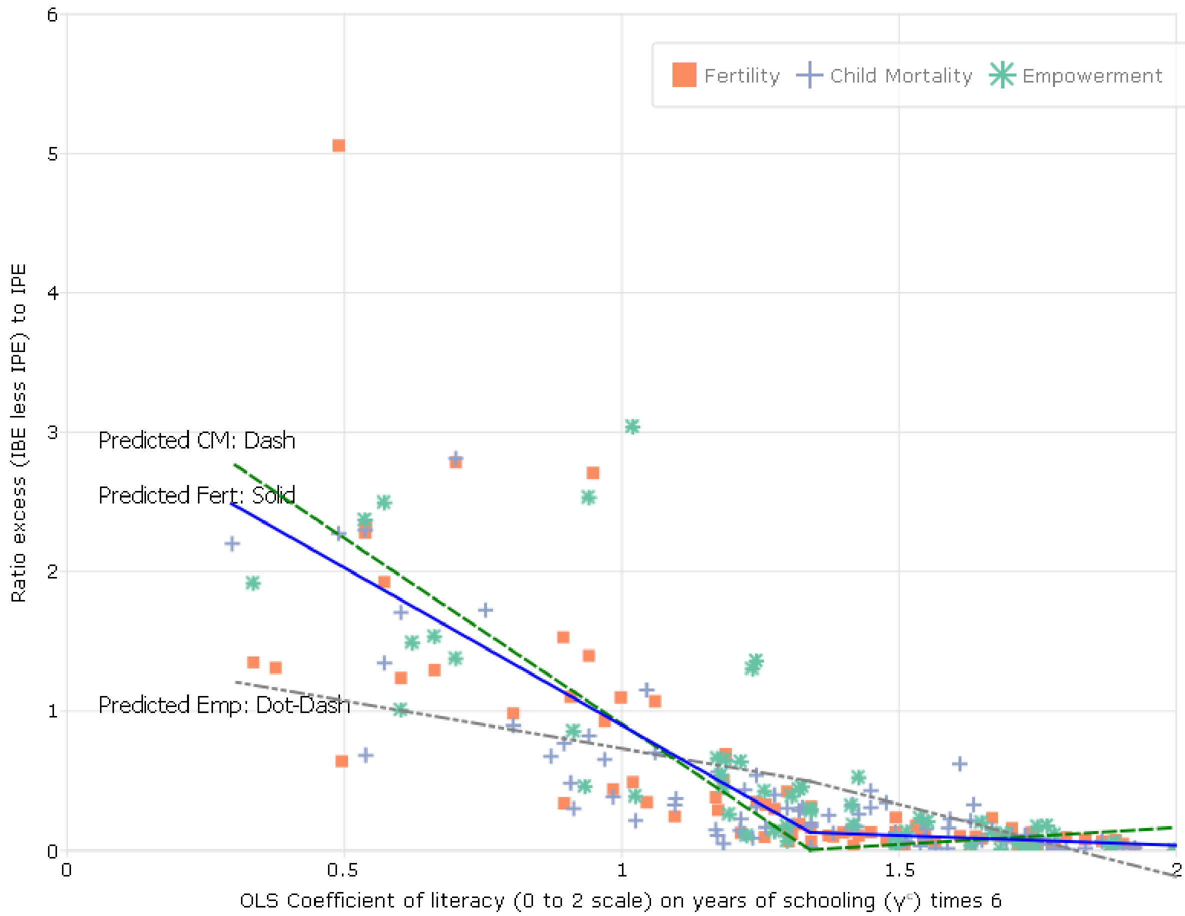
$$\text{Ratio of IBE-IPS over IPS} = \frac{\beta^c_{R(DHS)|S,Z} * (2 - \gamma^{c,R})}{\beta^c_{S|R,Z} + \beta^c_{R(DHS)|S,Z} * \gamma^{c,R}} \tag{12}$$

Equation 12 intuitively says that the excess of the estimated IBE over IPS depends on (i) how much of the impact on outcomes is due to reading versus the direct (partial) schooling effect and (ii) how large is reading acquisition in primary schooling ($\gamma^{c,R}$). For instance, if the partial impact of schooling is zero ($\beta^{c}_{S|R,Z} = 0$) then equation 12 reduces to $(2-\gamma^{c,R})/\gamma^{c,R}$ which tends to zero when primary schooling universally produces literacy (which is category 2 on the DHS literacy scales). If there is no direct effect of schooling, then the excess of IBE over IPS tends to infinity as $\gamma^{c,R}$ tends to zero, as if primary schooling produces zero literacy and all the causal channel is through literacy then IPS tends to zero.

In the estimates there is both a direct (parital) effect of schooling and an effect of reading. Figure 3 shows the scatter plot for each survey round of the excess impact of basic education over the impact of primary schooling against the OLS estimates of learning from six years of schooling for each of the three DHS outcomes. Intuitively, when primary schooling reliably produces learning (values of $\gamma^{c,R}$ above 1.5) then the gap is relatively small whereas in those countries where primary schooling tends not to reliably produce literacy (e.g. values of $\gamma^{c,R}$ less than 1) then the estimated impact of primary schooling is lower than the impact of basic education (schooling plus literacy). This is because, to the extent that the impact of schooling operates by creating a higher ability to read, this impact is lost when schooling does not produce learning. We saw in Figure 2 that the 10th percentile of $\gamma^{c,R}$ was .84 (with countries like Ghana and Nigeria well below even that low level). Hence in these countries the total impact of primary schooling will be much lower than the impact of basic education, which assumes reading is achieved.

Estimates that include a direct measure of learning provide some indication of the channels whereby schooling has its impact on outcomes, which, as shown above, is central to informed decisions about priorities for spending/investment/action/reform. While there are many, many estimates of the "returns to schooling" using pecuniary (e.g. wages or incomes) or non-pecuniary benefits *none* of these, even those estimates of the impact of schooling that are casually well-identified, are informative for choosing whether to invest in an additional

Figure 3: The impact of basic education on outcomes is larger than the impact of primary education, and much more so when the level of reading acquisition in primary schooling is low



Note: Only country/years where the literacy coefficient has the "right sign" are shown. The "predicted" lines are from a spline regression with a knot at the median of $\gamma = 1.34$ and is illustrative only. Three country/years (Egypt round 5, Mali round 7, Ghana round 5) with anomalous values are excluded.

year of schooling at existing learning or invest in raising learning. The estimated returns to schooling alone cannot decide this issue (even for private, much less public, spending).

We next move on to an estimation technique that allows us to adjust for measurement error. Given the sensitivity of the estimates of the relative pathways of schooling and literacy to measurement error we return to the implications of these estimates (for instance in MB/MC calculations) using the IV estimates.

5 Using Instrumental Variable (IV) Estimation Techniques to Adjust for Differential Measurement Error in Schooling and Literacy

Measurement error is a ubiquitous (and often severe) problem in all of econometrics, but our situation is the perfect storm of multicollinearity and differential measurement error for two reasons.

First, schooling and literacy are highly correlated (Appendix E has a brief technical primer). Hence measurement error in either variable will very strongly affect *both* estimates, making one too low (attenuation bias) and the other too high (as a consequence of what we call "partially omitted variable bias"). This makes estimates of the ratio of schooling and learning as causal channels (which, as we have seen, feeds into many formula, like the relative MB to MC of learning vs schooling in equation 10) doubly wrong (as discussed in Section 2).

Second, there are good reasons to believe that assessing whether a person can read one or a few arbitrary sentences or a passage is a very noisy measure of reading, and reading is a very noisy measure of literacy, and even a sophisticated measure of literacy is a very noisy proxy for the variety of learning results that potentially affect life outcomes. While schooling also suffers from measurement error, the measurement error in reading as a proxy for learning that affects life outcomes is likely much larger than errors in self-reported years (or level) of schooling. Differential *relative* measurement error is part of a perfect storm with highly correlated variables as the differential attenuation bias, which likely attenuates reading coefficients more than schooling, strongly affect OLS estimates of both terms.

5.1 Using Instrumental Variable Estimation Techniques to Correct for Measurement Error

We use instrumental variables estimation as a technique to correct for measurement error. To create instruments we take advantage of the clustered sampling used by both DHS and FII, in which respondents in the same enumeration area (EA) are geographic neighbors. We create a "enumeration area leave-out-mean" (EALOM) for each individual i, which is the average literacy (or schooling) level of everyone else in the individual's enumeration area j except individual i.

$$\bar{L}_{i,j} = \sum_{k=1}^{N_j, k \neq i} L_{i,j}/(N_j - 1); \bar{S}_{i,j} = \sum_{k=1}^{N_j, k \neq i} S_{i,j}/(N_j - 1) \tag{13}$$

where $L_{i,j}$ ($S_{i,j}$) is the literacy (schooling) of the i^{th} woman in the j^{th} EA and N_j is the total number of respondents in enumeration area j.

To produce consistent estimates an instrument must meet two criteria, first stage "inclusion" and structural equation "exclusion", and there are large literatures on "weak instruments" which demonstrate that the econometric consequences of failing to meet either of these two criteria are severe (Staiger and Stock (1997), Andrews et al. (2019)).

First, the "inclusion" criteria is that the instrument must be correlated with the variable being instrumented. Weakness in this condition leads to bias, imprecise IV estimates, and incorrect standard errors. A respondent's schooling and literacy levels are plausibly correlated with her sampling cluster neighbors' as they plausibly had similar opportunities for schooling attendance and may have attended similar quality schools. The F-statistics for inclusion of our EALOMs as instruments are typically above 10, a commonly used threshold for an adequate instrument[19]. However, as we estimate each survey/round separately, there is substantial variation across countries and we see instances in weak “first stage” instruments producing very imprecise and odd (e.g. wrong signed and excessively large (both positive and negative)) estimates.

The second criteria is that the instrument must satisfy the “exclusion” restriction: the instrument must not have a direct causal impact on the outcome of interest and can therefore be properly excluded from the equation of interest. There is no test for the exclusion restriction, and we cannot guarantee this for the EALOM instrument. If there are true “peer effects” or there are cluster specific factors affecting outcomes correlated with the instruments these would cause the exclusion criteria to be violated. As we only have one potential instrument (the “just identified” case) we have no method for testing these alternative hypotheses. The consequences of this are discussed in Appendix F, and further shown in Appendix G.

An alternative approach would be to run regressions at the EA level, rather than instrumenting with the EALOM. We run pooled regressions at the EA level (as well as at the woman and regional level) and report these results in Appendix H. The pooled, EA-level regression results move in the same direction as the IV results, showing larger relationships compared to OLS (indicating reduction of measurement error) for most coefficients, but to a lesser degree than IV.

5.2 Instrumental Variable Results

The results in the first column of Table 2 for each outcome are shown only to demonstrate how serious the perfect storm of measurement error with highly correlated variables is in this case. When only reading is instrumented the estimate of the association of reading with outcomes goes up substantially compared to OLS and hence, due to the multicollearity

[19] Stock and Yogo (2005) show that this threshold is not accurate and depends on a number of aspects of the problem. For a single variable to be instrumented and a single instrument the critical values range from 5.5 to 16.4 depending on the desired maximal size of a 5 percent Wald test (their Table 5.2).

of the estimates of schooling and reading, this drives the estimate of the direct (partial) impact of schooling down. The estimates of the partial schooling effect ($\beta_{S|R,Z}$) are large and of the "wrong sign" for three of the four outcome variables (fertility, child mortality, and empowerment) and essentially zero, for financial behaviors. That said, since this is a multi-collinearity problem the estimates of the impact of basic education are much more stable and reasonable as this is the linear combination of the two terms. But, any future empirical work attempting to disentangle the relative impacts of schooling and learning must cope with the differential measurement error and multi-collinearity problem as OLS, by not incorporating the differential attenuation bias will overstate the relative importance of schooling per se versus learning (as schooling has less measurement error and is less attenuated) whereas, as these results show, econometrically accommodating measurement error for only learning can produce results suggesting the direct (partial) impact of schooling is negative (which seems unlikely).

Our main focus is the second column of Table 2 for each outcome variables, which uses the EALOM instrument for both schooling and reading. These estimates have two important features.

First, the estimates of the impact of basic education on outcomes is much higher than OLS. The IV estimates for fertility suggest basic education is associated with a reduction of 1.24 children, from an average of 3.37. This is 3.3 times larger than the OLS estimate of 0.37. Basic education estimated with EALOM IV is associated with an increase in child survival of .077, which given that child survival in the sample was already 0.89, implies a two thirds reduction in child mortality from achieving basic education versus no schooling and no reading. Again this is almost three times higher than the OLS estimate of .027. Basic education increases the index of female empowerment of by 0.684 standard deviations, compared to just 0.19 using OLS. Basic education is associated with an increase in the financial behaviors index of 0.89 standard deviations, compared to an OLS estimate of 0.503.

Using instrumental variables techniques to account for the attenuation bias from measure-

Table 2: Method matters: Instrumenting for schooling and learning yields an estimated association of education (schooling + learning) with outcomes 3 times larger than that estimated by OLS

	Fertility (DHS)		Child survival (DHS)		Women's empowerment (DHS)		Financial Behaviors Index (FII)	
	EALOM IV, just reading	EALOM IV, Schooling and Reading	EALOM IV, just reading	EALOM IV, Schooling and Reading	EALOM IV, just reading	EALOM IV, Schooling and Reading	EALOM IV, just reading	EALOM IV, Schooling and Reading
Schooling (mean, RE meta-analysis weights)	0.971	-0.541	-0.047	0.029	-0.538	0.117	0.008	0.437
Std. Error of mean coefficient	0.126	0.121	0.011	0.01	0.093	0.108	0.1	0.093
20th percentile of results	-0.404	-2.359	-0.251	-0.078	-2.003	-0.804	-0.347	-0.686
80th percentile of results	3.39	0.967	0.096	0.154	0.42	1.451	0.43	0.758
Reading (mean, RE meta-analysis weights)	-2.01	-0.566	0.11	0.032	1.13	0.538	1.13	0.368
Std. Error of mean coefficient	0.194	0.119	0.018	0.011	0.168	0.128	0.168	0.158
20th percentile of results	-5.23	-2.27	-0.118	-0.078	-0.478	-1.004	0.026	-0.396
80th percentile of results	-0.118	0.905	0.452	0.22	3.97	1.525	1.596	2.416
Basic Education, linear combination of primary schooling and reading (mean, RE meta-analysis weights)	-0.832	-1.238	0.049	0.077	0.544	0.684	0.811	0.893
Std. Error of mean coefficient	0.053	0.088	0.005	0.007	0.058	0.084	0.11	0.128
20th percentile of results	-1.89	-2.542	-0.003	-0.002	0.026	-0.01	0.397	0.354
80th percentile of results	-0.24	-0.158	0.185	0.179	1.596	1.922	1.43	1.81
Ratio of reading to primary schooling		1.05		1.10		4.60		0.84
Ratio of IV estimate of basic education to OLS estimate		3.32		2.81		3.53		1.78
Number of survey rounds	128	128	128	128	67	67	10	10

Note: Regressions contain controls for age, age squared, age cubed, wealth index, a rural/urban dummy, and dummies for regions. Schooling coefficients have been scaled to reflect primary school (six years) completion; reading coefficients are scaled to reflect going from no reading to reading without help. "Basic education" is the sum.

ment error for both schooling and literacy suggests that the association of basic education with three of the four outcomes we examine is actually three times larger than the standard methods. These estimates suggest women's (and men's) education is *much* better for life outcomes than even was thought.

In spite of the issues with precision of estimation that are common with IV estimation, the t-statistics of meta-analysis estimate of the impact of basic education in the typical country are above 7 for all four outcome variables.

Second, it appears that the estimates of the impact of reading are much more affected by techniques that accommodate measurement error than are estimates of schooling[20]. This differential change in the estimates changes the ratio of basic education's impact coming from reading versus the direct (partial) schooling effect. Whereas with the OLS estimates the impact of reading was smaller than primary schooling (e.g. for fertility the scaled coefficient was 2.5 times larger for primary schooling) with the IV estimates they are roughly equal for fertility and child survival, slightly lower for financial behaviors but much larger for women's empowerment.

Figure 4 summarizes the primary results of the paper as is shows the distribution across all available survey rounds of the difference between the simple and standard approach to estimating the impact of women's schooling, which does not differentiate between schooling and education, and the returns to women's education which estimates jointly the completion

[20]The ratio of the IV to OLS estimates for fertility are 5.39 for reading while only 2.1 for the coefficient on schooling, for female empowerment the IV/OLS coefficient ratio is 4.98 for reading and 1.44 for schooling, for child survival these ratios 3.55 and 1.71. The financial behaviors index is the exception, where the ratio of IV to OLS is larger for schooling than for literacy. Recall from Appendix E equation 16 that the ratio of OLS to IV estimates (which is the inverse of the ratios above) is a crude, rough and ready, estimate of the signal to signal plus noise ratio. This suggests that the variability in the DHS literacy indicator is only about 20% signal (1/5.39) for the purpose of measuring the impact of learning on fertility, and 28% (1/3.55) for child survival. This might seen like an excessive degree of measurement error, but two points. One, ratios of OLS to IV estimates this small (and hence estimates of noise to signal this large) are not uncommon. Filmer and Pritchett (2001) show the ratio of OLS to IV estimates of the impact of household consumption per person on child school enrollment were .15 for Pakistan, .16 for Indonesia, and .46 for Nepal. Second, the measurement error is not the measurement error of literacy as a measure of literacy alone but also of this particular measure of literacy as a proxy for all other learning that may affect the outcomes. One can easily imagine the signal of reading a single, simple, passage is associated with, but only weakly, the extent to which all learning affects fertility choices or female empowerment.

Figure 4: Comparing the distribution of estimates of primary schooling with OLS versus basic education using IV

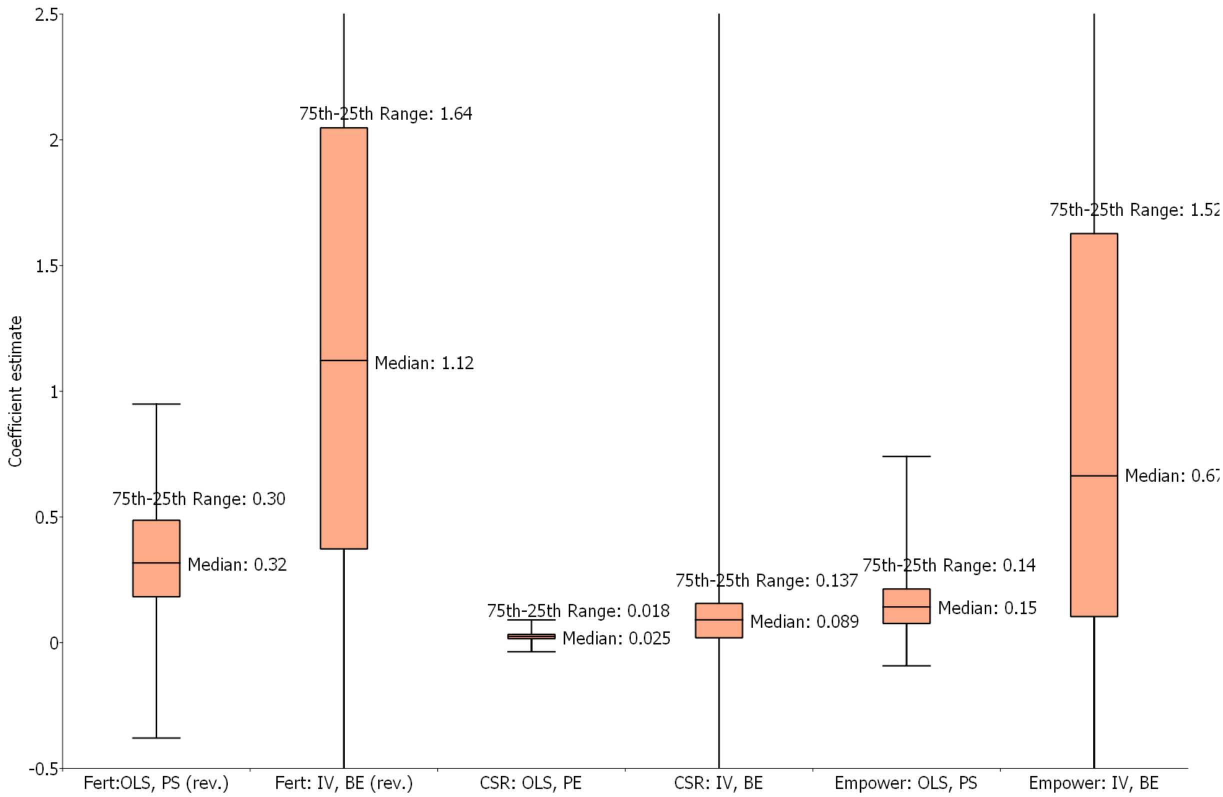


Table 3: Summary of methodological changes from basic OLS to IV to correct for measurement error in both schooling and literacy

	OLS, Schooling only	OLS, Schooling and Reading		IV, Instrumenting for Reading only		IV, Instrumenting for both Schooling and Reading	
	Schooling coefficient (proxying for "education")	Linear combination ("education")	Ratio to OLS with Schooling only	Linear combination ("education")	Ratio to OLS with Schooling only	Linear combination ("education")	Ratio to OLS with Schooling only
Dependent variable: Fertility (DHS)	-0.329	-0.373	1.134	-0.832	2.528	-1.238	3.762
	0.018	0.018		0.053		0.088	
Dependent variable: Child mortality (DHS)	0.023	0.027	1.189	0.049	2.122	0.077	3.339
	0.001	0.001		0.005		0.007	
Dependent variable: Women's empowerment (DHS)	0.146	0.193	1.324	0.544	3.729	0.684	4.686
	0.013	0.015		0.058		0.084	
Dependent variable: Financial Behaviors Index (FII)	0.279	0.503	1.803	0.811	2.908	0.893	3.201
	0.030	0.037		0.110		0.128	

Random effects meta-analysis estimates from 139 surveys in 54 countries. Regressions contain controls for age, age squared, age cubed, wealth (using the a wealth index), a rural/urban dummy, and dummies for regions. Schooling coefficients have been scaled to reflect primary school completion; literacy coefficients have been scaled to reflect going from illiterate to literate..

of primary school and acquiring at least literacy, combined with using methods that adjust estimates for measurement error. The *typical* (median) of the distribution of the estimates is much higher for each of the outcomes. On the other hand, the downside of using instrumental variables is that the estimates in each individual country (survey-round) are much less precise and hence the dispersion, in these box plots measured as the 75th-25th range is also much larger.

Table 3 shows the set of estimates for each life outcome for each of the estimation variants to show how much difference each of the specification and method changes makes. The move from OLS to IV estimation accounts for most of the rise in the estimated impact of basic education.

5.3 Forward looking caveats

Before examining the implications of these estimates and concluding we would like to highlight three major limitations of our results.

First, as is discussed and shown with comparisons of the empirical cumulative distribution functions of the OLS and IV results in Appendix F and Appendix G, the imprecision of the IV (standard error bounds) becomes very large in many cases. That the estimates of the impact of basic education combine the estimates of schooling and literacy further complicates things. Because the coefficient estimates are so highly correlated, in many instances one of the coefficient estimates is wildly implausible (large and of the wrong sign) but that is compensated in the linear combination by the other coefficient being implausibly large in the opposite direction so that the linear combination estimate of basic education is of plausible magnitude even when neither of the individual terms is. This implies that the decomposition into schooling and learning is plausible only in aggregation, not necessarily case by case.

We have been consistent that although we use the word “impact” for convenience we are not claiming to establish causality from observational data. This raises the second and third limitations of our results, the second being generic and the third specific to our decomposition into schooling and learning.

The second limitation is the standard issue that the number of years of schooling a woman has is determined by choices that she (and, as a child, her family) make and that these choices might reflect characteristics that are not in our (paltry) set of controls. This would lead those who are more likely to have good outcomes to have acquired more schooling and hence the observational association of schooling and outcomes would overstate the causal impact of a given person receiving more schooling.

There are two things about our study that are different from the setting in which this has been most explored, the wage returns to schooling. One, by examining non-pecuniary outcomes (not money wages in employment) we mitigate the implications of signaling as there is not “third party” to whom more schooling is a signal. Two, as suggested above, the fact that the DHS sample is censored above–only those with primary school as their highest level are included in the DHS sample–this means that we are only comparing outcomes among those women who did not choose (or were not able) to attend secondary school or higher. This raises the possibility at least that much of the observed variation across women in whether they attended primary school or not was due to factors like physical access that might have been "as if" it were exogenous.

To be clear, our results are no worse in this lack of causal identification and lack of robustness than nearly all of the existing literature estimating the returns to women's schooling as there are so few studies with plausible identification, Mensch et al. (2019) for instance, find in their systematic review only 16 studies on the question of women's education and child and maternal health that pass their filters for identification.

The third issue is specific to the decomposition of pathways into schooling and learning. The intuitive answer to "how bad is the bias from lack of random assignment?" depends how much of the variation in the independent variables in the data is "as if" it were due to random assignment in that whatever determined the value of the independent variable was not correlated with the outcome. One can imagine that historically lots of people lived in rural areas, schools were relatively rare, and what schools there were were "as if" randomly placed relative to the characteristics of the people who attended them. In such a scenario, whether or not the adult women we observe in the DHS have no schooling or have primary schooling might heavily (mostly?) depend on whether there was a (somewhat randomly placed) school nearby when they were young. In this case the bias from selectivity, that women who had primary schooling also have characteristics more likely to make them have good outcomes, might be modest[21]. However, in order to identify the literacy (or more generally, the learning) impacts one needs variation in the amount of measured learning of individuals with the same degree of schooling. While some of that variation may be "as if" randomly assigned because some children were proximate to good (high value added) schools and others happened to be proximate to bad (low value added) schools, the evidence is pretty powerful that far and away the most powerful correlates with measured learning

[21] One could argue that the most striking thing about the famous Duflo (2001) use of the rapid expansion of schools in Indonesia in the 1970s as an exogenous "natural experiment" to instrument for schooling attainment to estimate wage returns to schooling is how very much like the simplest OLS the sophisticated IV estimate is.

are child background characteristics (like SES). And, it is quite easy to believe that the same individual characteristics that account for higher learning, conditional on attending a given level of schooling are those characteristics that lead to more favorable life outcomes. The upshot of this is that there are reasons to believe the bias in OLS coefficients relative to a LATE are larger for learning than for schooling.

We label these "forward looking" caveats as we are engaged in trying to address these issues and hope in future work to make progress on them.

6 Illustration of costs and benefits of expanding schooling versus raising learning

With the OLS and IV estimates of the coefficients of schooling and literacy in producing beneficial life outcomes we can return to equation 10 which gives, for any given values of S, γ, $\beta_{S|L,Z}$ and $\beta_{L|S,Z}$ the highest cost an action that improves learning can be and still be optimal, relative to the cost of increasing schooling (normalized to 1). Table 4 illustrates, with actual magnitudes, the implications of equation 10. It uses unscaled coefficients $\beta_{S|R,Z}$ and $\beta_{R|S,Z}$ from Tables 1 and 2.

The first implication is, the lower the level of learning currently produced by a year of schooling (γ) the higher the relative costs that would be optimal to incur to improve learning. For example, when $S = 6$, using the IV coefficients (columns 1-3), the benefit in terms of outcomes of improving learning by one unit relative to increasing schooling by one year (the benefit of which is normalized to 1) is 18 at low learning versus only 10.5 when learning is high.

Second, the relative benefit of improving learning is higher when the level of schooling is higher, so this is 19.8 when $S = 9$ (column 7) and only 6.6 when $S = 3$ (column 8). This is intuitive as the higher learning per year applies to more years of schooling. Countries that have already achieved relatively high levels of schooling attainment but at low levels

Table 4: The highest optimal cost of increasing learning relative to schooling (cost of achieving one additional year set to 1) given the estimated pathways to outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Schooling years	6	6	6	6	6	6
Estimation Technique	IV	IV	IV	OLS	OLS	OLS
Literacy gain per year of schooling (γ)	0.11	0.22	0.33	0.11	0.22	0.33
Fertility	14	11.1	9.3	6.5	5.8	5.2
Child Survival	14.5	11.5	9.5	8.1	7.1	6.3
Women's Empowerment	32.9	20.5	14.9	16.6	12.7	10.3
Financial Behaviors Index	11.9	9.7	8.3	20.4	14.8	11.7
Average	18.3	13.2	10.5	12.9	10.1	8.4

Note: This is equation 10 using unscaled coefficients from Tables 1 and 2. The cost of increasing school attainment by one year is normalized to 1.

of learning could vastly increase life benefits by increasing the learning per year from their schooling.

Third, the relative benefits of investing in learning are higher the larger the relative channel of impact on outcomes is through learning versus the "direct" effect of schooling. This can be seen in two ways. As seen in Table 2 the relative impact of learning to schooling is higher for empowerment (0.538 versus 0.117) than for financial behaviors (0.368 versus 0.467) and hence (in Table 4) at $S = 6$ and $\gamma = .22$ the relative benefit of learning to schooling is 20.5 for empowerment but only 9.7 for financial behaviors. The other way is that the OLS estimates, which estimate that less of the relative channel of impact is through learning than does IV, produce consistently lower ratios for each outcome than do the IV (columns 4-6).

So the choice of investing in improving learning versus expanding schooling (beyond what countries consider a basic right–we are in no way suggesting anything less than universal primary completion is an acceptable policy or goal) depends on relative costs. In low learning environments it may be that investments in improving learning are *orders of magnitude* more cost effective than spending that expands attendance in low learning schools.

7 Conclusion

This paper empirically illustrates the practical importance of two conceptual points and we do this using four different outcome variables across two completely different cross-nationally comparable data sets.

The first conceptual point is that if the causal pathway whereby schooling affects outcomes is (even partially) through learning (skills, competencies, capabilities, acquired factual knowledge, analytic and reasoning ability) then, to the extent that schooling in different settings is more or less effective in conveying learning, the impact of *schooling* and the impact of *education* (schooling plus the learning it is intended to convey) are not conceptually the same. Any estimate of the impact of schooling will *under*estimate the impact of education to the extent that schooling does not produce the intended learning. We empirically illustrate this point by showing two things.

One, both in the DHS and FII data the observed learning profiles (increase in literacy per year/level of schooling) are (a) much less than one might expect (e.g. only half of women with primary schooling complete can read, even at a rudimentary level) and (b) widely varied across countries (Figure 2).

Two, we empirically illustrate the implications of this conceptual point with standard OLS estimates of the association of four non-pecuniary outcomes with schooling and learning in Table 1 and in Figure 3. We show that for "typical" estimated coefficients (the random effect weighted averages) the association of basic education is higher than that of primary schooling by 13 to 80 percent (Table 1) and that in a low learning environment the association of basic education with outcomes can be twice as high as that of primary schooling (at observed learning levels) (Figure 3 based on equation 12). So, while an enormous literature demonstrates the beneficial impacts of *women's schooling* the benefits of *women's education* are even larger, even using OLS estimates.

Moreover, if one compares the instrumental variable (IV) estimates, which correct for differential measurement error using enumeration area leave out means, to OLS estimates of women's schooling (the state of play in the existing literature) these estimates are three to four times larger. Women's (and men's) education is much better in IV estimates than we thought from OLS estimates.

The second conceptual point is that the policy implications for investing in education of estimates of the benefits of education on life outcomes depend on the causal pathways whereby schooling produces benefits, and any calculation of the "optimal" spending allocation depends on the marginal cost and marginal benefits (via the causal pathways) of alternatives. Even the simplest formula for the optimal pattern of spending between, say, expanding schooling and increasing learning while in school as a means of producing better life outcomes for women in any given country (equation 10) depends on: the current level of schooling (S^c), the current level of learning (γ^c), the "direct" (non-learning mediated) impact of schooling on outcomes $\beta^c_{S|L,Z}$ and the impact of learning on outcomes ($\beta^c_{L|S,Z}$). The standard illustration of the benefits of schooling to justify more "investment" in education is valid, but lacks specificity to guide the choice of investments and the pattern of spending.

We empirically illustrate this point with both OLS and IV estimates, showing that, given the implications of differential measurement error for the estimation of the *relative* importance of the two channels (equations 16 and 19), method matters, a lot. The IV estimates suggest that when measurement error is addressed, for three of the four outcomes the learning pathway is much stronger than the "direct" schooling pathway. Table 4 shows the implications for the cost of an intervention that raises learning versus expanding schooling, for various levels of S and learning (γ).

We emphasize that we "empirically illustrate" these conceptual points, for reasons both positive and to allow ourselves needed caveats.

On the plus side, we illustrate these conceptual points not for one outcome in one country but with four different outcomes, using two completely different data sets, for over 50 different countries. So our results are truly illustrative of both the "typical" country and the variability across countries. We start from the illustration with the simplest cross-tabulation of outcomes by level of schooling to illustrate that our results empirically encompass, and then extend, a huge literature on the association of schooling, and particularly women's schooling, and a variety of outcomes as our IV regressions are just a (modest) advance in method; the same conceptual points can be made with the simplest cross-tabulations.

We use the term "empirically illustrate" also however to acknowledge the many weaknesses of our existing results relative to the ideal of consistent/unbiased estimates of causal impacts, while at the same time emphasizing that the weaknesses of the empirical work do not affect the conceptual points. In particular, two points. One, we make no claim that our estimates of any of what we in this paper, strictly for convenience, called "impacts" actually represent well identified estimates of the causal LATE. That said, even a perfect experiment that produced a causally well identified and precise estimate of the impact of primary schooling would not (a) estimate the impact of basic education, (b) have external validity because, among other reasons, variation in learning per year of school implies the impact of schooling will differ if learning matters, or (c) inform the allocation of investment across expanding years of completed schooling versus learning per year of schooling as the relative magnitude of the pathways matters. Two, our IV estimates using enumeration area leave out means (EALOM) produce results that are interesting and intriguing but which cannot be regarded as definitive, which does not in any case suggest the OLS results are more reliable, only that getting the estimation of the relative casual pathways done in a convincing way in the face of the perfect storm of differential measurement error across strongly correlated variables is a high priority for future research.

8 Bibliography

Lant Pritchett. *The Rebirth of Education: Schooling Ain't Learning*. Center for Global Development, 2013. ISBN 978-1-933286-77-8.

Michelle Kaffenberger and Lant Pritchett. More Schooling or More Learning? Evidence from Learning Profiles from the Financial Inclusion Insights Data. *RISE Working Paper*, 17/012, 2017.

Lant Pritchett and Justin Sandefur. Girls' Schooling and Women's Literacy: Schooling Targets Alone Won't Reach Learning Goals. *Center for Global Development Policy Paper*, 104, 2017. doi: 10.35489/BSG-RISE-WP_2017/011.

Esther Duflo, Pascaline Dupas, and Michael Kremer. The Impact of Free Secondary Education: Experimental Evidence from Ghana. page 105, 2019.

Mari Oye, Lant Pritchett, and Justin Sandefur. Girls' schooling is good, girls' schooling with learning is better. *The Education Commission Background Paper*, 2016.

J. G. Cleland and J. K. Van Ginneken. Maternal education and child survival in developing countries: The search for pathways of influence. *Social Science & Medicine (1982)*, 27 (12):1357–1368, 1988. doi: 10.1016/0277-9536(88)90201-8.

Sarah Cochrane. The effects of education on health. *World Bank Sector Working Paper*, SWP405:1, 1980.

United Nations. *Socio-Economic Differentials in Child Mortality in Developing Countries.* UN,, 1985. ISBN 978-92-1-151154-3.

Emmanuela Gakidou, Krycia Cowling, Rafael Lozano, and Christopher J. L. Murray. Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: A systematic analysis. *Lancet (London, England)*, 376(9745):959–974, 2010. doi: 10.1016/S0140-6736(10)61257-3.

Teresa Castro Martin. Women's Education and Fertility: Results from 26 Demographic and Health Surveys. *Studies in Family Planning*, 26(4):187–202, 1995. doi: 10.2307/2137845.

Emily C Keats, Anthony Ngugi, William Macharia, Nadia Akseer, Emma Nelima Khaemba, Zaid Bhatti, Arjumand Rizvi, John Tole, and Zulfiqar A Bhutta. Progress and priorities for reproductive, maternal, newborn, and child health in Kenya: A Countdown to 2015 country case study. *The Lancet Global Health*, 5(8):e782–e795, 2017. doi: 10.1016/S2214-109X(17)30246-2.

Barbara S. Mensch, Erica K. Chuang, Andrea J. Melnikas, and Stephanie R. Psaki. Evidence for causal links between education and maternal and child health: Systematic review. *Tropical Medicine & International Health*, 24(5):504–522, 2019. doi: 10.1111/tmi.13218.

Stephanie R. Psaki, Erica K. Chuang, Andrea J. Melnikas, David B. Wilson, and Barbara S. Mensch. Causal effects of education on sexual and reproductive health in low and middle-income countries: A systematic review and meta-analysis. *SSM - Population Health*, 8: 100386, 2019. doi: 10.1016/j.ssmph.2019.100386.

Lucia Breierova and Esther Duflo. The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers? *National Bureau of Economic Research*, 10513, 2004. doi: 10.3386/w10513.

Una Okonkwo Osili and Bridget Terry Long. Does female schooling reduce fertility? Evidence from Nigeria. *Journal of Development Economics*, 87(1):57–75, 2008. doi: 10.1016/j.jdeveco.2007.10.003.

Anthony Keats. Women's schooling, fertility, and child health outcomes: Evidence from Uganda's free primary education program. *Journal of Development Economics*, 135:142–159, 2018. doi: 10.1016/j.jdeveco.2018.07.002.

Iqbal Dhaliwal, Esther Duflo, Rachel Glennerster, and Caitlin Tulloch. Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. 12 2013. URL http://hdl.handle.net/1721.1/116111.

Alejandro J. Ganimian and Richard J. Murnane. Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, 86(3): 719–755, 2016. ISSN 00346543, 19351046. URL http://www.jstor.org/stable/24752880.

Paul Glewwe and Karthik Muralidharan. *Improving Education Outcomes in Developing Countries*, volume 5 of *Handbook of the Economics of Education*, chapter 10, pages 653–743. Elsevier, 2016. doi: 10.1016/B978-0-444-63459-. URL https://ideas.repec.org/h/eee/educhp/v5y2016icp653-743.html.

Elizabeth Denison Brown and Jeffery Tanner. Integrating value for money and impact evaluations : Issues, institutions, and opportunities. *World Bank Policy Research Working Paper*, WPS 9041, 2019. URL http://documents.worldbank.org/curated/en/862091571145787913/Integrating-Value-for-Money-and-Impact-Evaluations-Issues-Institutions-and-Opportunities.

Felipe Barrera-Osorio, Andreas de Barros, and Deon Filmer. Long-term impacts of alternative approaches to increase schooling: Evidence from a scholarship program in cambodia. *World Bank Policy Research Working Paper*, No. 8566, 2018.

OECD. PIAAC Literacy: A Conceptual Framework. *OECD Education Working Papers*, 34, 2009. doi: 10.1787/220348414075.

Sunita Kishor and Lekha Subaiya. Understanding women's empowerment: A comparative analysis of demographic and health surveys (DHS) data. 2008.

Douglas Staiger and James H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586, May 1997. URL https://ideas.repec.org/a/ecm/emetrp/v65y1997i3p557-586.html.

Isiah Andrews, James Stock, and L. Sun. Weak instruments in iv regression: Theory and practice. *Annual Review of Economics*, 11:727–753, 2019.

James Stock and Motohiro Yogo. *Testing for Weak Instruments in Linear IV Regression*, volume Identification and Inference for Econometric Models, chapter 5, pages 80–108. New York: Cambridge University Press, 2005.

Deon Filmer and Lant H. Pritchett. Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. *Demography*, 38(1): 115–132, 2001. doi: 10.2307/3088292.

Esther Duflo. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4):795–813, 2001. doi: 10.1257/aer.91.4.795.

Paul Glewwe. Why Does Mother's Schooling Raise Child Health in Developing Countries? Evidence from Morocco. *The Journal of Human Resources*, 34(1):124–159, 1999. doi: 10.2307/146305.

Kirk Dearden, Lant Pritchett, and Jeff Brown. Learning From Neighbors: Social Learning About Child Feeding During Diarrheal Episodes. 2011.

Joshua D. Angrist. The perils of peer effects. *Labour Economics*, 30:98–108, 2014. doi: 10.1016/j.labeco.2014.05.008.

A Omitted variables bias

All estimates from observational data are subject to omitted variable bias. Suppose that for any given outcome, Y, there is the set $\mathbf{L}$ of all the N_L measures of learning, $\mathbf{L} = (L_1, L_2, ...L_{n_L})$ and the set $\mathbf{Z}$ of all the other non-schooling or learning covariates that affect outcome Y (and suppose the relationship is truly linear) then the "all else" in equation 1 could be an individual specific idiosyncratic measurement error that is orthogonal to S, $\mathbf{L}$, $\mathbf{Z}$ and hence standard estimation techniques, like OLS, would produce estimates that were consistent for (converge in the limit to) their conceptual counter-parts. However, any actual data set gives us a very paltry set of available L and Z. The $L_{included}$ and $Z_{included}$ are a small sub-set of the true $\mathbf{L}$ and $\mathbf{Z}$[22]. Using only $L_{included}$ and $Z_{included}$, the "all else" in equation 1 includes the $L_{excluded}$ and $Z_{excluded}$ variables that are, by assumption, correlated with Y and potentially correlated with S. Hence a standard OLS estimate of $\beta_{S|L_{included}, Z_{included}}$ does not converge to $\beta_{S|\mathbf{L},\mathbf{Z}}$ and the direction of the magnitude and direction of this "omitted variable" bias depends on the pattern of associations between Y, S, and the $L_{excluded}$ and $Z_{excluded}$.

In the special case of only L and no Zs the formula for omitted variables bias says the OLS regression coefficient of Y on S alone (with L omitted) would converge in probability limit to:

$$\beta^c_{S|L=\emptyset} \rightarrow \beta^c_{S|L} + \gamma^c * \beta^c_{L|S} \tag{14}$$

where γ^c is the "true" impact of S on L. Hence the OLS estimate of schooling on outcomes excluding any learning measure in equation 14 is very similar to the expression for the *total* impact of primary schooling in equation 4 which includes the pathway of impact of S through L. Hence while it will not be exact (as we have ignored the Zs and other determinants of learning besides S), the simple OLS regression of outcomes on S (omitting L) is a rough and ready estimate of the *total* impact of schooling, or the partial impact of schooling plus its impact through the pathway of the learning it produces.

[22] Bivariate cross-tabs, correlations, and regressions using only Y and S are just a special case, $L_{included} = \emptyset$, $Z_{included} = \emptyset$

Any estimate of the "impact of schooling" not conditioned on L will *overestimate* the *direct* impact of schooling conditional on learning (the partial derivative) if (as expected) S and L have a positive correlation. And such estimates of the impact of schooling will *underestimate* the impact of basic education (equation 5) to the extent that less is learned in primary school that what is considered basic education ($\Delta L > \gamma^c * \Delta S$), as is illustrated in Figure 1, and the magnitude of this bias depends on pace of learning, which is expected to be different in each country, as shown in Figure 2.

B Comparing measured literacy: DHS and FII

Appendix Table B.1 shows literacy levels among women who completed primary school from the FII surveys compared to women from the same 10 countries who completed grade 6 from the DHS. The results are reassuringly close on average and in correlation. The estimated literacy level of women with primary schooling/six years complete is 51% for the DHS and 49% for the FII. The correlation across the two sources is 0.77. Though there are some countries that are substantially different (e.g. the DHS suggest very low literacy in Ghana while the FII show literacy in Ghana is about average. The DHS suggests very high literacy in Rwanda whereas the FII is higher than average, but not as far above the mean).

Table B.1: Assessed ability of women with just primary schooling to read a simple sentence or passage is similar between the DHS and FII data

Country	DHS, women 25-34, highest grade was 6th, percent able to read all of a sentence	FII survey, women aged 18-37, completed primary, able to read a sentence
Nigeria	12.0%	15.4%
Uganda	54.4%	23.2%
Bangladesh	32.6%	29.5%
Pakistan	50.7%	44.2%
India	34.6%	49.0%
Kenya	65.3%	69.7%
Indonesia	75.2%	76.7%
Tanzania	86.2%	82.5%
Ghana	7.7%	47.9%
Rwanda	97.1%	77.7%
Average	51.4%	48.8%

Source: Pritchett and Sandefur (2017), and authors' calculations with FII data.

C Tables comparing the summary statistics of the estimated results for fertility, child survival and women's empowerment

Table C.1: Summary statistics for estimated results (OLS and IV) of association of primary schooling, reading, and basic education with fertility

	Primary Schooling			Reading		Basic Education (primary schooling plus reading)	
Summary statistics of estimates	OLS schooling alone	OLS both	IV both	OLS both	IV both	OLS both	IV both
Median	-0.317	-0.282	-0.592	-0.1	-0.596	-0.357	-1.122
Average	-0.331	-0.262	-0.324	-0.109	-1.138	-0.371	-1.461
Percent right signed	93.80%	85.20%	67.20%	69.50%	64.10%	95.30%	87.50%
Number positive and significant	2	5	10	2	5	1	0
Number negative and significant	100	79	29	40	26	105	60
20th percentile	-0.536	-0.473	-2.356	-0.257	-2.27	-0.562	-2.542
80th percentile	-0.147	-0.044	0.967	0.05	0.89	-0.199	-0.247
Standard deviation	0.229	0.251	6.142	0.176	6.627	0.228	2.47
Random effects meta-estimate	-0.331	-0.26	-0.538	-0.106	-0.567	-0.374	-1.238
Random effects standard error	0.018	0.02	0.121	0.015	0.119	0.018	0.087
Absolute value ratio RE estimate to std err	18.5	12.9	4.5	7.2	4.8	21.1	14.2
Number of survey round estimates	128	128	128	128	128	128	128

Summary statistics for the estimates across all survey rounds for fertility.

Table C.2: Summary statistics for estimated results (OLS and IV) of association of primary schooling, reading, and basic education with child survival rate

	Primary Schooling			Reading		Basic Education (primary schooling plus reading)	
Summary statistics of estimates	OLS schooling alone	OLS both	IV both	OLS both	IV both	OLS both	IV both
Median	0.025	0.017	0.041	0.009	0.042	0.03	0.091
Average	0.024	0.017	0.057	0.01	0.093	0.028	0.15
Percent right signed	91.40%	85.20%	64.80%	75.80%	58.60%	92.20%	80.50%
Number positive and significant	92	60	15	29	11	92	49
Number negative and significant	1	1	4	3	1	1	1
20th percentile	0.013	0.004	-0.078	-0.001	-0.077	0.016	0.003
80th percentile	0.035	0.031	0.15	0.022	0.22	0.043	0.179
Standard deviation	0.017	0.016	0.286	0.019	0.484	0.02	0.431
Random effects meta-estimate	0.023	0.017	0.029	0.009	0.032	0.027	0.077
Random effects standard error	0.001	0.001	0.01	0.001	0.011	0.001	0.007
Absolute value ratio RE estimate to std err	18.3	12.8	3	7.9	3	20	10.6
Number of survey round estimates	128	128	128	128	128	128	128

Summary statistics for the estimates across all survey rounds for child survival.

Table C.3: Summary statistics for estimated results (OLS and IV) of association of primary schooling, reading, and basic education with women's empowerment

Summary statistics of estimates	Primary Schooling			Reading		Basic Education (primary schooling plus reading)	
	OLS schooling alone	OLS both	IV both	OLS both	IV both	OLS both	IV both
Median	0.142	0.071	0.246	0.096	0.499	0.183	0.664
Average	0.162	0.092	0.498	0.121	0.586	0.213	1.085
Percent right signed	86.40%	78.80%	54.50%	89.40%	68.20%	93.90%	81.80%
Number positive and significant	44	25	11	30	16	50	27
Number negative and significant	1	2	8	1	4	0	0
20th percentile	0.062	0	-0.769	0.028	-0.878	0.095	0.075
80th percentile	0.237	0.171	1.222	0.209	1.467	0.307	1.895
Standard deviation	0.149	0.157	2.985	0.133	3.087	0.17	2.756
Random effects meta-estimate	0.147	0.081	0.117	0.109	0.538	0.195	0.684
Random effects standard error	0.013	0.014	0.108	0.013	0.128	0.015	0.084
Absolute value ratio RE estimate to std err	11.2	5.8	1.1	8.3	4.2	13.3	8.1
Number of survey round estimates	66	66	66	66	66	66	66

Summary statistics for the estimates across all survey rounds for women's empowerment. There are 66 survey rounds because Sierra Leone round 5 was excluded as its IV results failed.

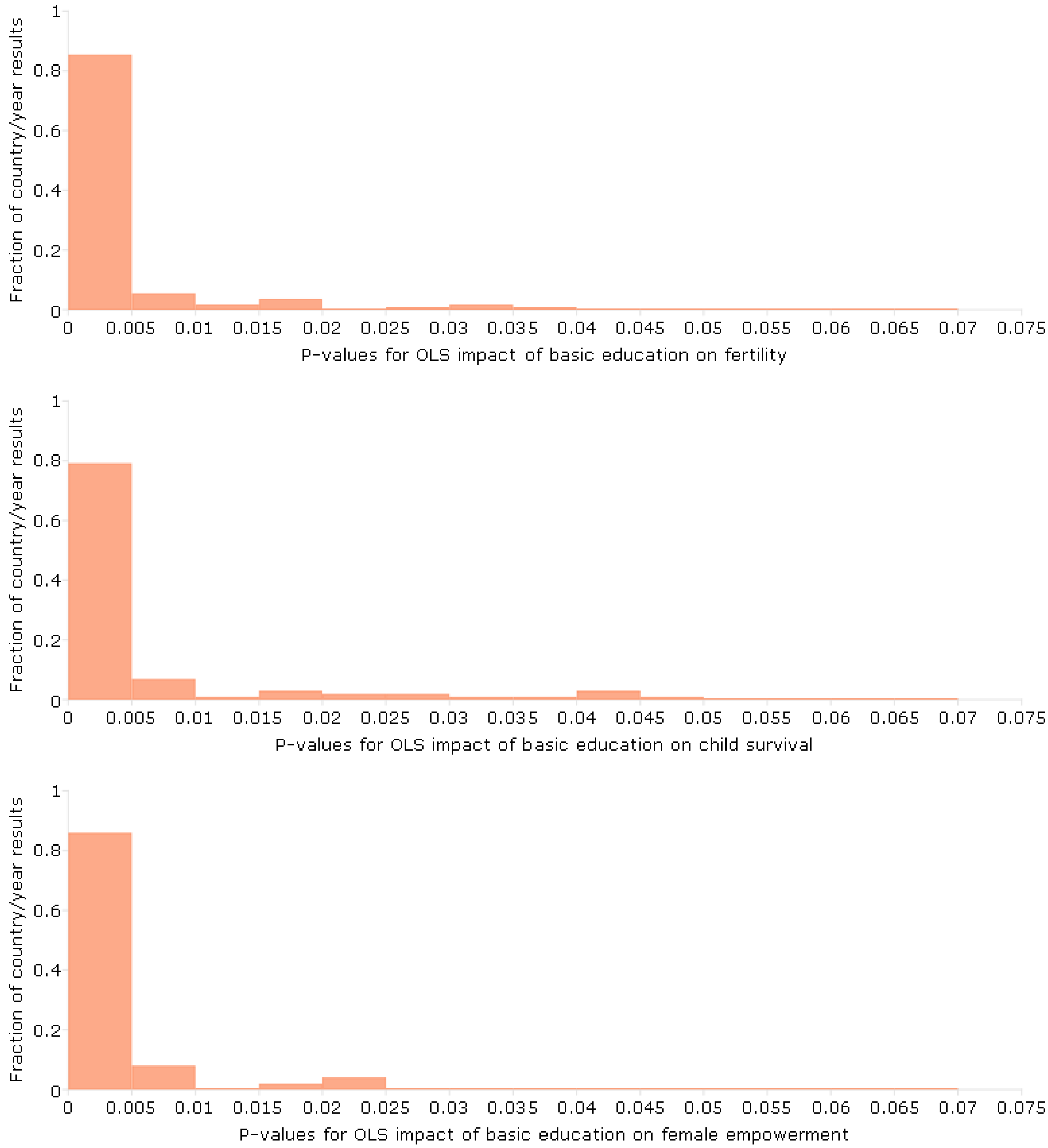
D Graphs showing the lack of p-hacking using DHS data sets

While observational data from cross-nationally comparable data sets like the DHS nearly always lack the necessary ingredients to create clean causal identification, there is a big advantage in repeating exactly the same empirical analysis across multiple data sets. This avoids the dangers of data-mining in a study of a particular topic or country (in which variable definitions, functional forms, specifications, etc. are chosen to produce "good" results). This also avoids "p-hacking" that can affect attempts at systematic reviews of the literature as, if there is a bias towards the publication of papers/studies that show results that are "statistically significant" at some threshold p-level (like the common "5 percent"), then reviews of the published literature will include studies that have too much clustering around those p-levels.

This can extend to the non peer-reviewed working papers if authors only write up and put into working paper form results that are "statistically significant". Alternatively, a major problem in reviewing literatures is what the paper is written, titled, and abstracted to be "about" may depend on which, of the many possible, coefficients turned out to be significant. These forms of p-hacking can easily affect even "systematic reviews" that rely on standards for "rigor" that include the credibility of the approach to identification as, since registries of studies are relatively new and announced pre-analysis plans still relatively rare in the social science, the combination of data and "definition" mining (even after experiments), publication, write up, and "aboutness" bias can still create a tendency towards published studies with positive findings just above some arbitrary p-level.

As can be seen from Figure D.1, the histogram of p-levels across our studies changes around the .05 (or any other) conventional level of statistical significance are completely absent in our OLS results of the estimates of the impact of basic education (the linear combination of six years of schooling and achieving literacy). This is in part because, given the sample size of each DHS survey/year and the statistical strength of the relationship the vast majority of the p-levels are under .005 percent, much less than conventional "t-test" levels of .01/.05/.10.

Figure D.1: P-hacking graphs for OLS results on basic education (schooling plus literacy)



E Measurement error and attenuation bias

Measures in survey data contain measurement error. Our simple measures of literacy contain measurement error of multiple types. First, we are using one single question/passage to assess an entire domain of "ability to read" and hence perhaps a woman could have read other sentences but not the particular one(s) presented in the DHS/FII surveys. Second, the literacy tests are administered by enumerators who must judge how well a respondent could or could not read and these judgments are almost certainly not perfectly consistent across enumerators. Moreover, despite strenuous efforts at quality control in these surveys, enumerators could also mis-record or even just skip the question and fill in an arbitrary answer. Third, literacy is only one aspect of learning and any of the outcomes (e.g. child mortality, fertility) could be affected by learning mathematics, learning actual relevant facts (e.g. about child health, see Glewwe (1999)), or by schooling acquired skills (Mensch et al., 2019). Fourth, the literacy measures have few categories and are sharply right censored (women who read very well have the same measure of literacy of those who can barely read the sentence).

Measures of schooling could also contain measurement error if years of schooling is mis-reported by respondents, or mis-recorded by enumerators.

Measurement errors produce attenuation bias. Even in simple cross-tabulations of experience of a child death by levels of schooling and literacy (Figure 1) the differences across categories are attenuated if women with schooling are mis-classified as not having literacy when they do (or vice versa).

Equations 15 and 16 illustrate the simplest possible bivariate case with OLS if outcomes are related just to learning, and the literacy measure for a given woman is the "true" learning variable, L, plus an error term as in equation 15. The magnitude of the attenuation bias is a ratio of the signal to signal plus noise in the variable as in equation 16. This is intuitive as if this ratio is 1 there is zero noise ($\sigma^2_\omega = 0$) and hence no bias, while if there is no signal at all in literacy as a proxy for the true relevant learning variable then attenuation bias is complete and the estimated coefficient is the association with measurement error and will tend to zero ($\sigma_\omega^2 \to \infty,\ \beta_{OLS} \xrightarrow{p} 0$)).

$$\text{Measurement Error} : Literacy_i = L_i^* + \omega_i \tag{15}$$

$$\text{Attenuation bias} : \beta_L^{OLS} \xrightarrow{p} \beta_L * \frac{\sigma_L^2}{\sigma_\omega^2 + \sigma_L^2} \tag{16}$$

But what happens if there are two variables and one or both has measurement error and the degree of measurement error differs between the two variables?

The exact multivariate OLS estimate with two variables can be estimated with repeated OLS, so that if the true model is:

$$Y = \beta_S * S + \beta_L * L + \epsilon_i \tag{17}$$

then we can regress Y on L and S on L and regress the residuals of those regressions and recover numerically exactly the coefficient estimate on S from a multivariate regression of Y on S and L:

$$(y_i - \hat{\beta}_L * L_i) = \beta_S * (S_i - \hat{\pi}_L * L_i) + \eta_i \tag{18}$$

Where "$\hat{\beta}$" is the multivariate OLS estimate.

This is the intuition behind the formula for omitted variables bias, as the simple bivariate regression of Y on S is the equivalent of the above procedure but where β_L and π_L are forced to equal 0 rather than their actual OLS values. Think of estimating the following equation:

$$(y_i - \tilde{\beta}_L * L_i) = \beta_S * (S_i - \tilde{\pi}_L * L_i) + \eta_i \tag{19}$$

where the *tilde* represents any arbitrary value. If $0 < \tilde{\beta}_L < \beta_L$ then the estimate of S will suffer from what we call "partial omitted variables bias." Suppose that L is included in the regression but measured with error and that S is not. Repeated least squares will not produce an estimate for S that converges to the true value but will, if π_L is positive, be *overestimated*. When L is measured with error, attenuation bias is the same as if *part* of its effect on Y is omitted, and with omitted variables bias, positive correlation will imply *some* of that impact is attributed to S. The conclusion is that OLS multivariate regression of Y on S and L when L has measurement error and S and L are correlated will produce estimates on S ($\beta_{S|L}$) that are too large and estimates on L ($\beta_{L|S}$) that are too small. The consequences of differential measurement error for estimates that decompose the total impact of education on outcomes through the schooling conditional on learning and learning conditional on schooling as in equation 20 can be severe.

$$\text{Fraction of impact of education due to learning} = \frac{\beta_{L|S} * \Delta L}{\beta_{L|S} * \Delta L + \beta_{S|L} * \Delta S} \tag{20}$$

Suppose, at the extreme, that the true value of the "direct" impact of schooling was zero and hence the *entire* impact of education was through learning ($\beta_{S|L} = 0, \beta_{L|S} > 0$) so the true fraction in equation 20 is 1. But suppose the the ratio of noise to noise plus signal for the measure of learning was .5 and the correlation of S and L was .9. Then the estimated fraction of the impact of education due to learning would be only .065 even though we know in this hypothetical case the true fraction is 1.

In a situation where both schooling and learning are measured with error, if the simple literacy measure has a much larger degree of measurement error as a proxy for learning than does reported completed years of schooling has for completed years of schooling then the above will still be true. Schooling will be overestimated, and learning will be underestimated if the measurement error is not addressed.

We attempt to address this, in Section 5 using instrumental variables to address measurement error. While the IV approach has very real limitations, it illustrates the crucial importance of dealing with attenuation bias when attempting to decompose the associations or impacts of schooling and learning.

F Limitations of EALOM ("enumeration area leave out means") as Instrumental Variables

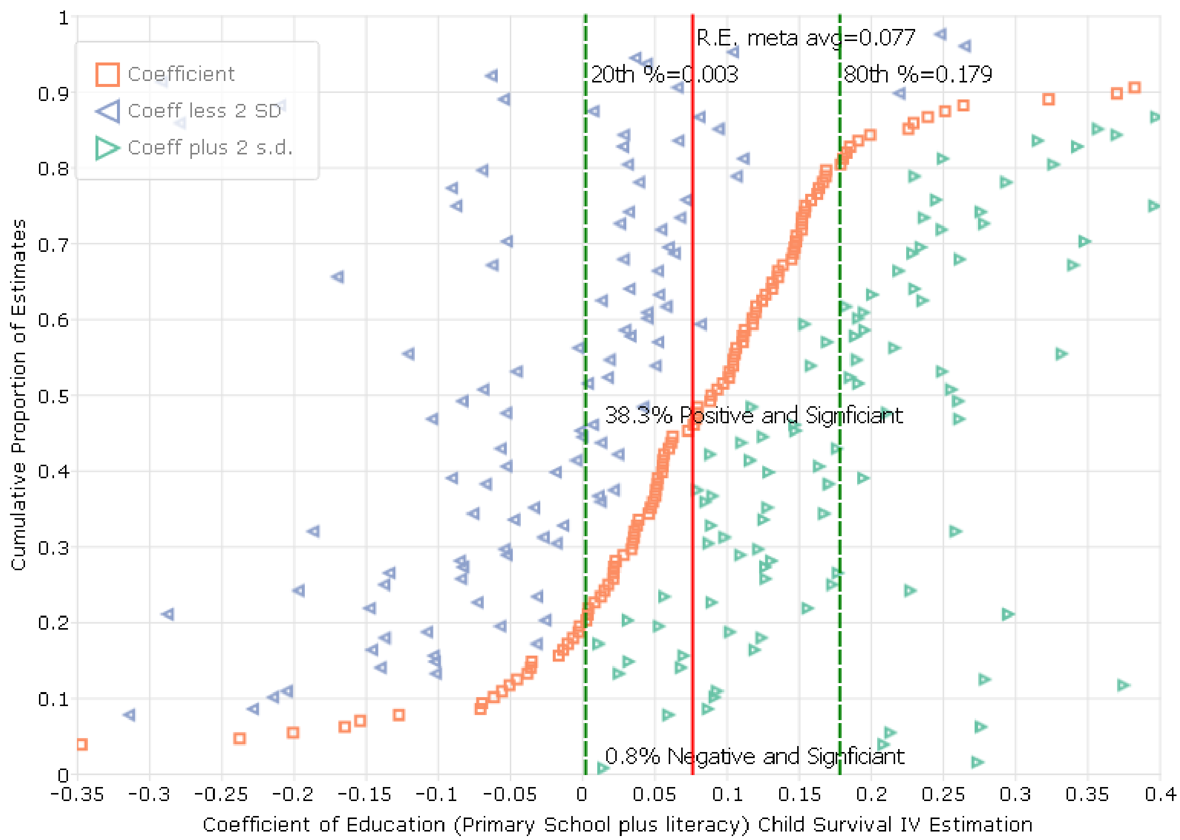
The instrumental variables estimates using enumeration area leave out means (EALOM) of literacy come with two major limitations.

First, while the estimated effects are much larger using instrumental variables, the precision of estimation is lower in each country and the variation in the estimates of the coefficient on literacy increases across countries as well. Furthermore, as explained above, since the covariance of the estimates for schooling and literacy is so high, the estimates of the ratio of pure schooling to literacy effects becomes very unstable. Figure F.1 shows the empirical cumulative distribution function of the IV estimates of the impact of basic education on child survival and the two standard error confidence intervals around those estimates (similar figures for the other outcome variables are in Appendix G). As in Table 2 the random effects weighted average of the IV estimates of the impact of education are much higher than their OLS counterparts (the weighted average is .077 vs .027). But Figure F.1 shows that the variance of the IV estimates in each country and the variability across countries is large. While the t-test for the null hypothesis that the random effects mean of the IV estimates of the impact of education is zero is around 10, this is due to the benefits of aggregation (and this is half that of the OLS estimate, 23, due to increased imprecision in estimation), the variability of estimates for each country is large. The 20th percentile of the estimates is essentially 0 and the 80th percentile is .179.

Figure F.2 shows the empirical cumulative distribution function for the 128 estimates of the impact of schooling (β_S) and impact of basic education estimated with OLS and with IV for child survival (these same graphs for fertility and female empowerment are in Appendix G.[23] As we saw in Section 4 the impact of education (estimated with OLS or IV)

[23]In this graph each of the estimates is sorted from lowest to highest and each "row" of the graph is the nth largest estimate and hence these are not for the same country. That is the 10^{th} largest estimate for OLS schooling, OLS education and IV education can be three different countries.

Figure F.1: Empirical cumulative distribution functions of the IV estimates of the impact of basic education on child survival with standard error bounds



Note: EALOM IV estimates of basic education (the linear combination of the scaled coefficients on schooling and literacy) and respective standard errors from 128 DHS survey rounds in 54 different countries.

is consistently higher than the impact of schooling by an amount that depends on learning.

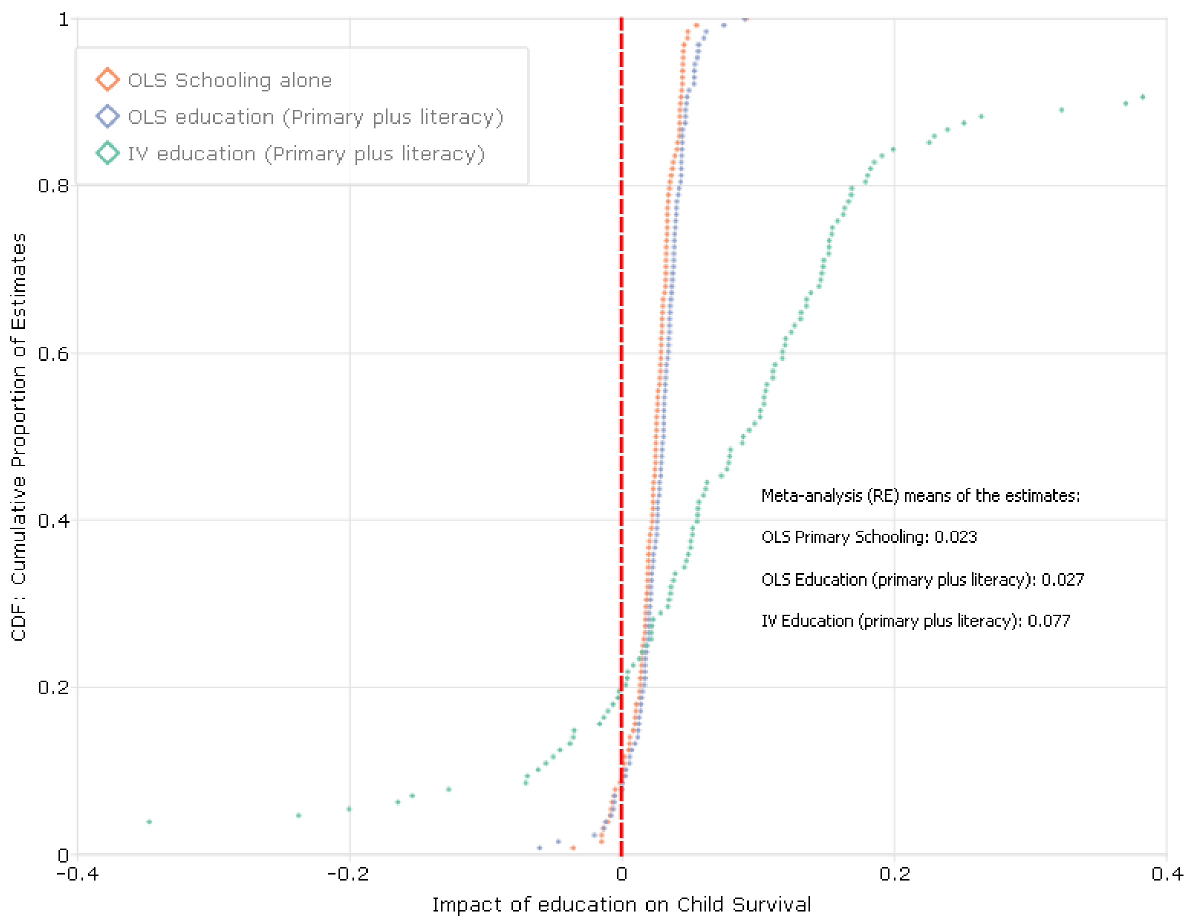
Figure F.3 compares, for each country, the empirical cdf of the IV results and the OLS results[24], and shows the IV estimate less 2 standard errors and hence the lower end of the confidence interval that an individual IV estimate is greater than zero. This graph shows that, while the average of the IV estimates is much higher there are many countries where the IV estimate is less than the OLS estimate. In fact, nearly 20% of IV estimates are negative. The greater imprecision implies that, while for the OLS estimates 93 of 114 are positive and statistically significant, for IV only 43 of 114 are.

A related cost of the much higher imprecision of the IV estimates is that this imprecision, combined with multicollinearity and hence high covariance of the estimates of schooling and literacy is that the decomposition of the total impact of education into the schooling channel and the literacy channel is unreliable country by country. Given the high covariance, when the estimated impact of literacy is high this induces a low estimate of schooling and hence the estimates of schooling and literacy are negatively correlated across countries for each outcome. This includes driving the estimated effects to "wrong-signed" values - 24% of the IV estimates of the partial effect of schooling ($\beta_{S|L,Z}$) on child survival are negative and 31% of the estimates for fertility are positive, both of which are intuitively implausible. Any precision or reliability of the estimates of the relative impacts of literacy and schooling is the result of aggregation across many countries/periods that smooths over this IV-technique-induced variability in the individual terms.

The second major issue with the IV estimates is whether the EALOM satisfies the exclusion restriction. It cannot be ruled out that the level of literacy of other women in a woman's neighborhood (enumeration area) *directly* affects her own outcomes. Perhaps having more literate neighbors leads to women having better health information as they learn from their neighbors (for some weak evidence to this effect see Dearden et al. (2011)), or, perhaps living in a neighborhood with more literate women, who themselves have fewer children, causes

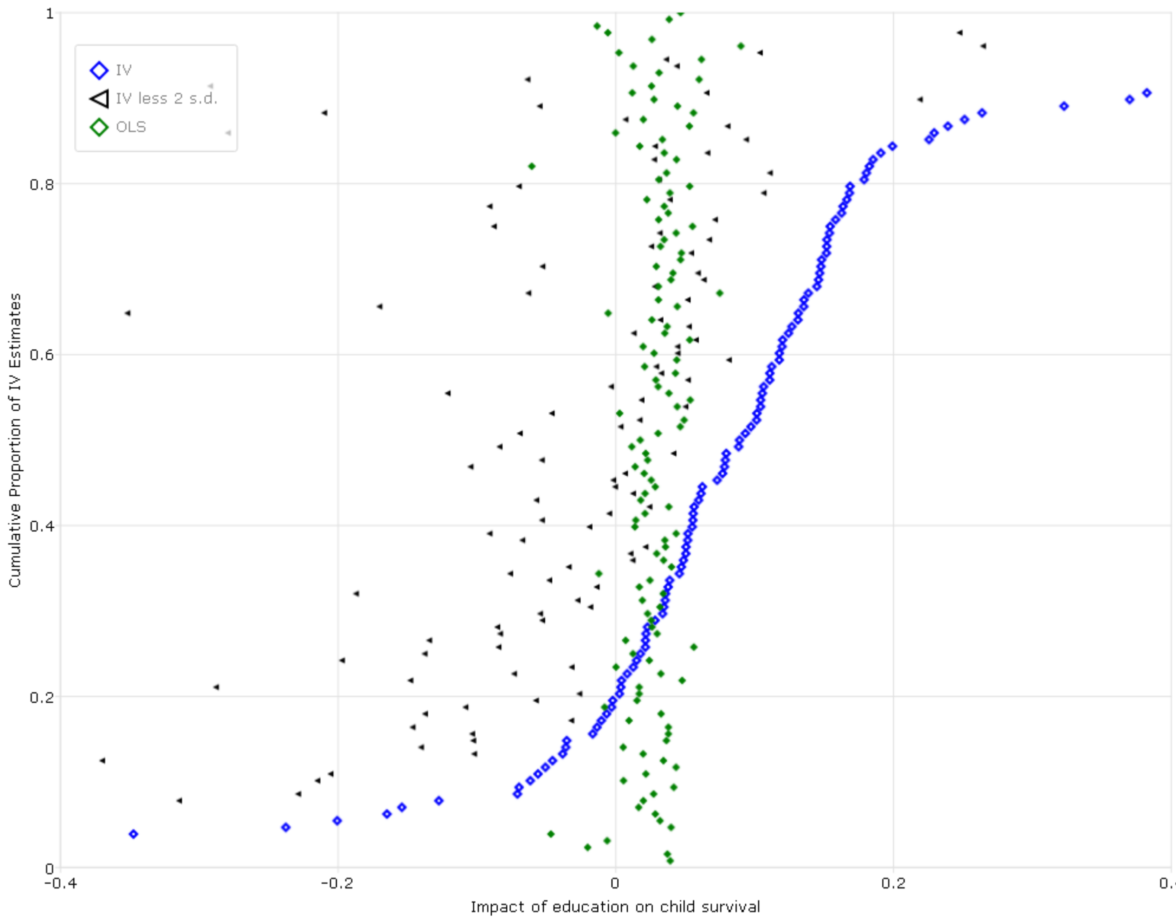
[24]Hence, in contrast to Figure F.2, here each "row" on the vertical axis is a country and they are sorted by the magntude of the IV estimate.

Figure F.2: Empirical cumulative distribution functions of the estimates of the impact of schooling (OLS) and basic education (OLS and IV) for child survival



Note: For 128 country/year survey rounds from 54 countries the empirical cumulative distribution function for (a) OLS with "schooling alone", the scaled coefficient from bivariate regression of outcomes on years of schooling, (b)The OLS estimate of the "impact of basic education" which is the linear combination of the scaled coefficients of schooling and literacy, (c) the EALOM IV estimates of the "impact of basic education." Each set of estimates is sorted and hence each horizontal row (the nth estimate) is not from the same survey year.

Figure F.3: The empirical CDF of the EALOM IV and OLS estimates of the impact of basic education on child survival, with IV 2 std error bounds



Note: (a) EALOM IV estimates of basic education (the linear combination of the scaled coefficients on schooling and literacy), (b) the IV estimates less two std errors, (c) the OLS estimate, all across 128 DHS survey rounds in 54 countries. Both the OLS and IV are sorted by the IV estimate so each horizontal "row" is a survey/year.

women to reassess their own preferences for children or living in a neighborhood with more empowered women leads directly to more empowerment for a woman. Moreover, any of a large number of possible neighborhood (enumeration area) effects cannot be ruled out and if these EA fixed effects are correlated with the EA level of literacy then the exclusion restriction is also (indirectly) not met, or, more particularly, the use of IV does not solve, and may exacerbate this omitted variables problem.

Two points. There is literature suggesting that true peer effects are often overstated (Angrist, 2014) as many findings of peer effects are just the result of the exclusion of local variables and it is nearly impossible to disentangle "locality" effects and "true" peer effects.

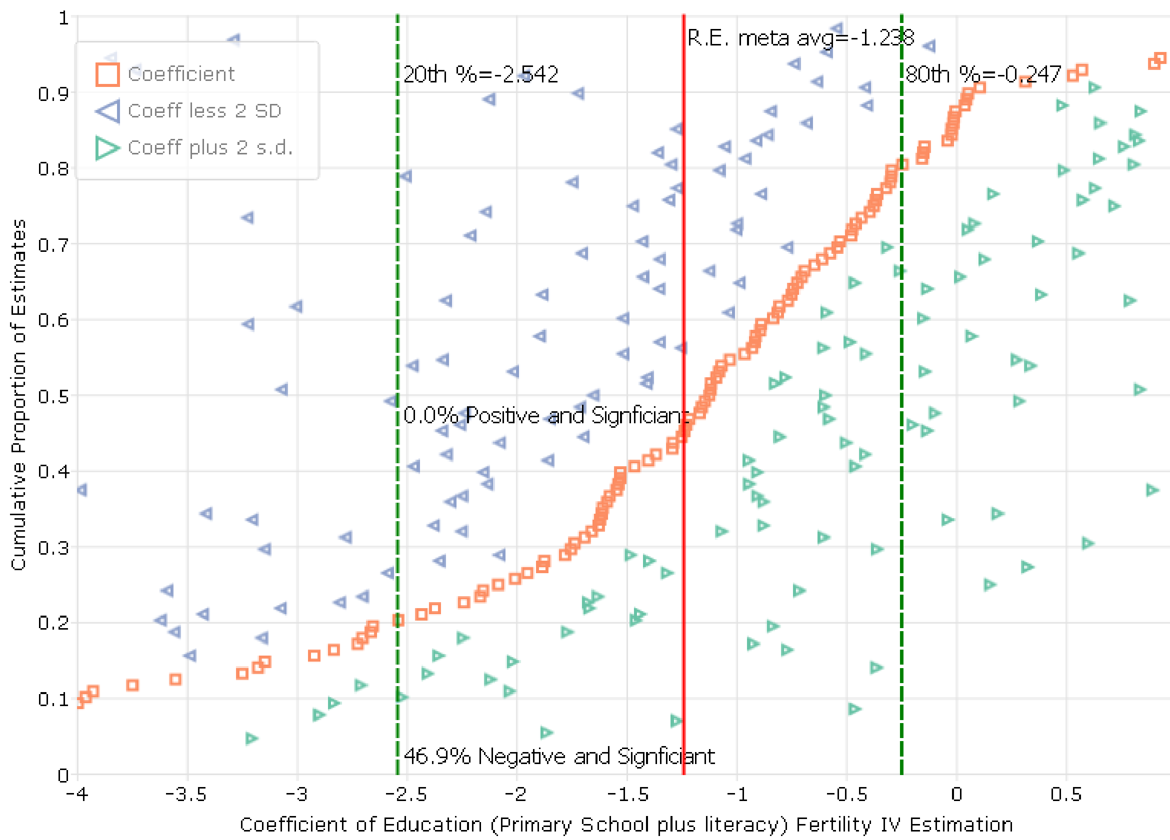
One test would be to look at how results from the original IV regressions differ between rural and urban areas. It is conceivable that if peer effects exist they would be stronger in rural areas – rural communities are likely to be smaller, with individuals knowing each other and interacting on a more regular basis. In urban areas which are more densely populated, residents are less likely to interact with and know a large proportion of fellow enumeration area residents and therefore peer effects would be diminished.

We test this with the FII data, running the IV regressions separately for rural and urban residents. It shows that, with the exception of Uganda, in all countries where the instrument is valid (F>10, six of the eight countries), it is valid for both the urban and rural subsamples, and in all countries where the instrument is not valid (F<10, three countries), it is not valid for either urban or rural subsamples. Only in Uganda is the instrument valid in rural but not urban areas. Further, the incremental R2s are mixed, with the instrument adding more in the rural regression in some countries, and more in the urban regression in other countries.

But more deeply, if there are true peer effects then this makes our IV estimates of the direct effect of a woman's literacy on her own outcomes "biased" but this would mean the total aggregate effect of improving women's education on outcomes is actually higher than just the sum of the effects for individual women. If there are true peer effects then there are positive externalities of education (of at least some geographic scope) and the impact on say, child survival, of increasing the literacy of one woman is the impact on her own outcomes plus the sum of the impact she has on all her connected peers. So if our IV estimates of the direct effect is "too high" because of peer effects, the direct effect is "too low" as an estimate of the total impact of increased education.

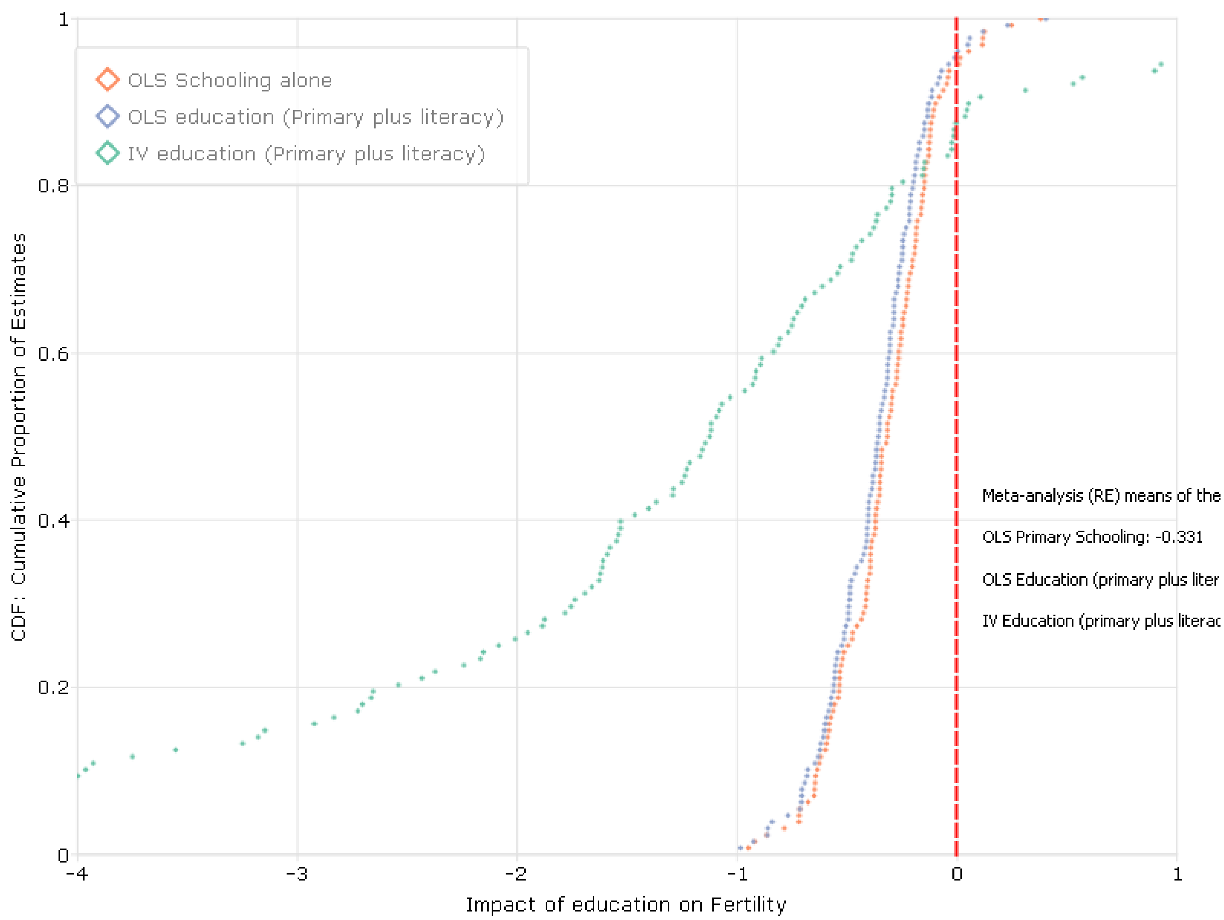
G The empirical cumulative distribution functions of OLS and IV estimates for fertility and empowerment

Figure G.1: Empirical cumulative distribution functions of the IV estimates of the impact of basic education on fertility with standard error bounds



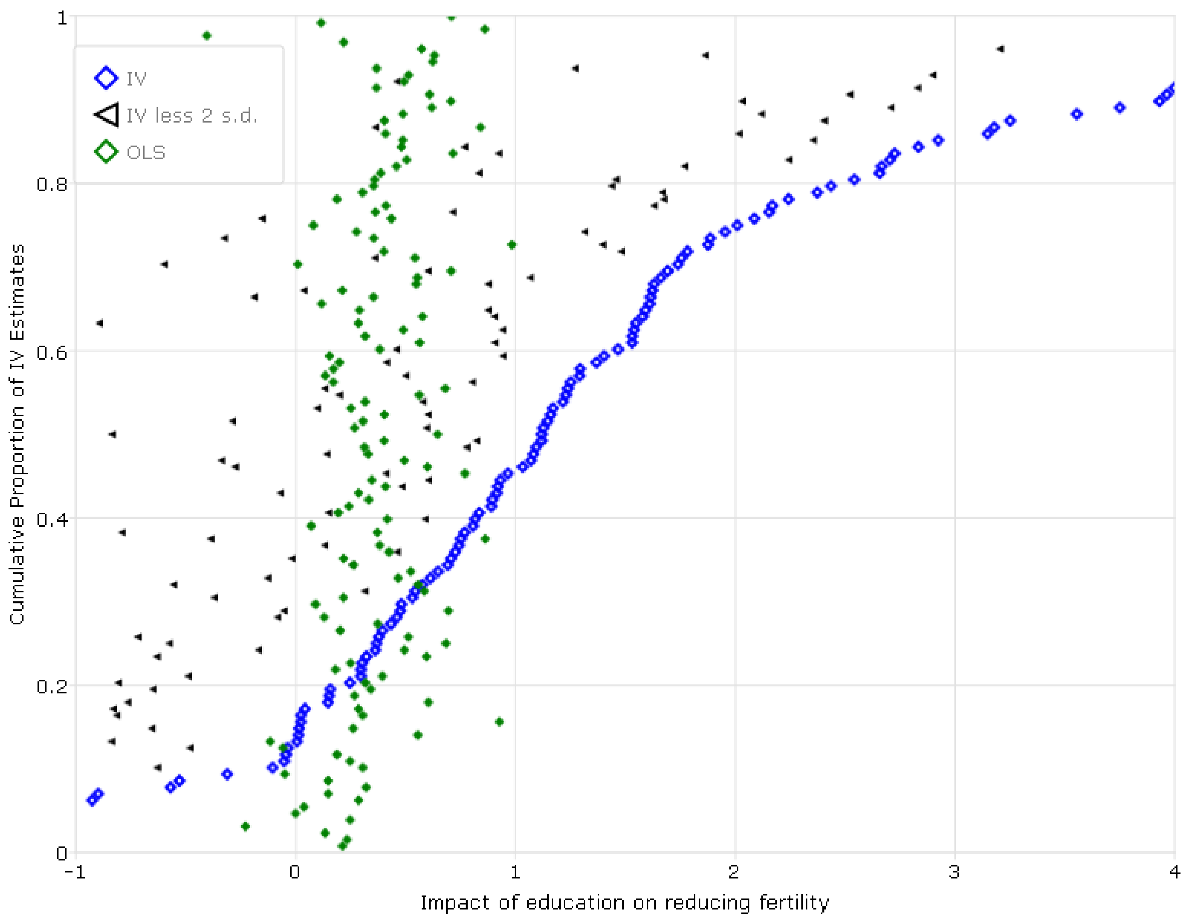
Note: EALOM IV estimates of basic education (the linear combination of the scaled coefficients on schooling and literacy) and respective standard errors from 128 DHS survey rounds in 54 different countries.

Figure G.2: Empirical cumulative distribution functions of the estimates of the impact of schooling (OLS) and basic education (OLS and IV) for fertility



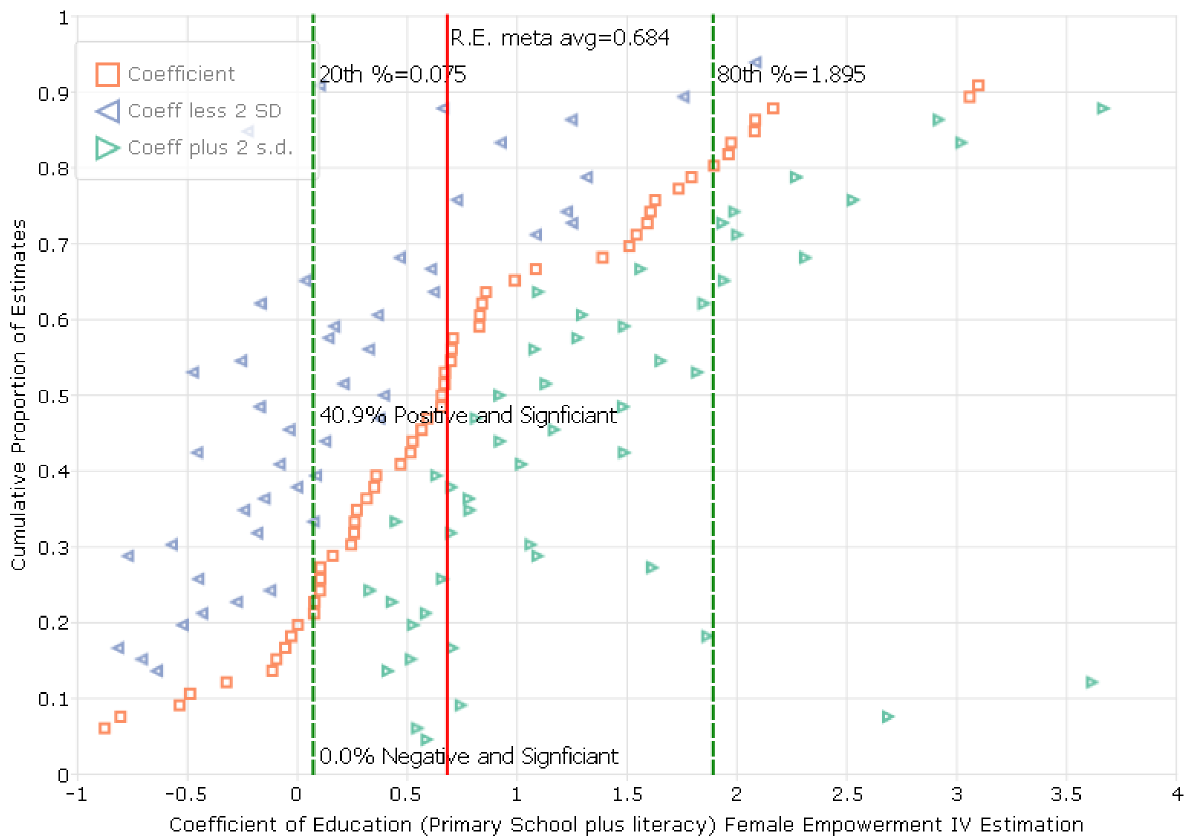
Note: For 128 country/year survey rounds from 54 countries the empirical cumulative distribution function for (a) OLS with "schooling alone", the scaled coefficient from bivariate regression of outcomes on years of schooling, (b) the OLS estimate of the "impact of basic education" which is the linear combination of the scaled coefficients of schooling and literacy, (c) the EALOM IV estimates of the "impact of basic education." Each set of estimates is sorted and hence each horizontal row (the nth estimate) is not from the same survey year.

Figure G.3: The empirical CDF of the EALOM IV and OLS estimates of the impact of basic education on fertility, with IV 2 std error boundss



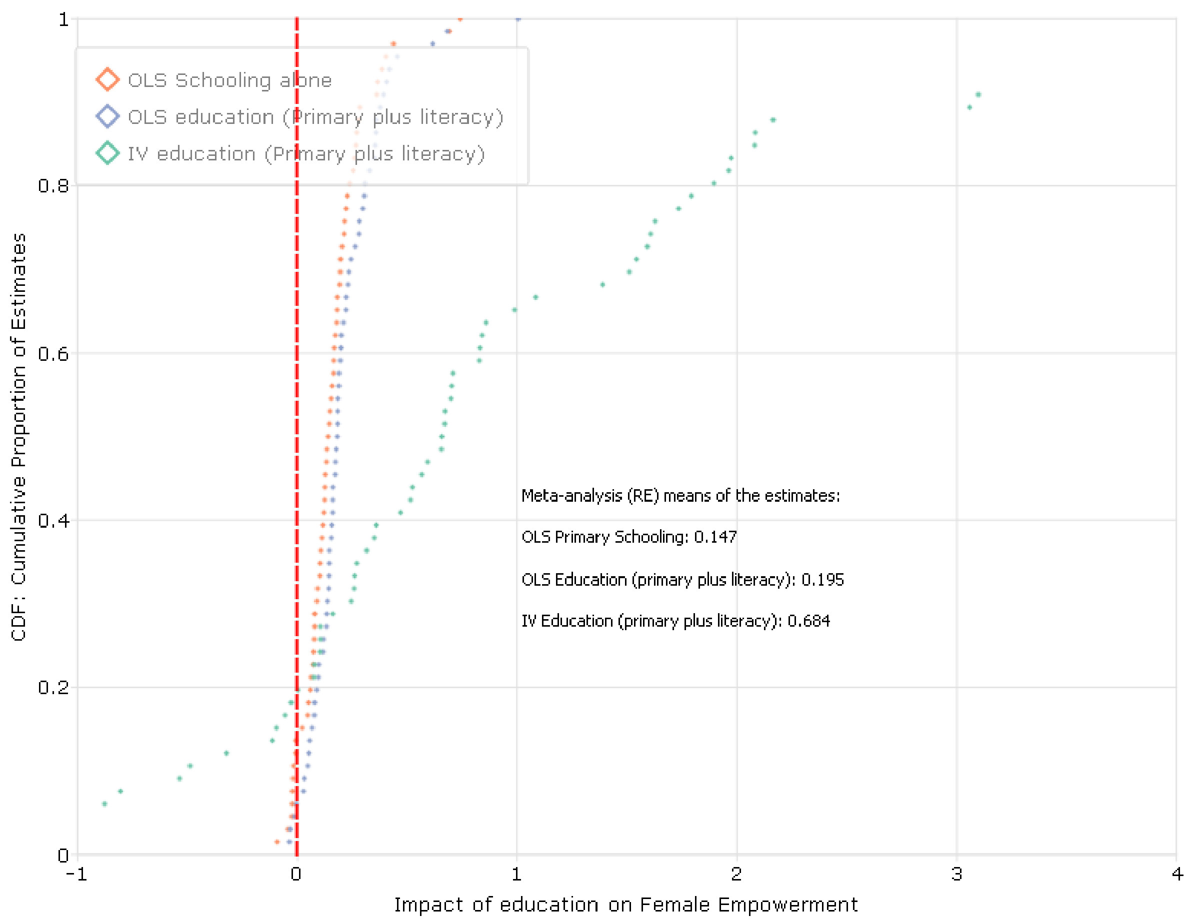
Note: (a) EALOM IV estimates of the sum of the scaled coefficients on schooling and literacy (and respective standard errors), (b) the IV estimates less two std errors, (c) the OLS estimate across from 128 DHS survey rounds in 54 different countries. Both the OLS and IV are sorted by the IV estimate so each horizontal "row" is a survey/year.

Figure G.4: Empirical cumulative distribution functions of the IV estimates of the impact of basic education on women's empowerment, with standard error bounds



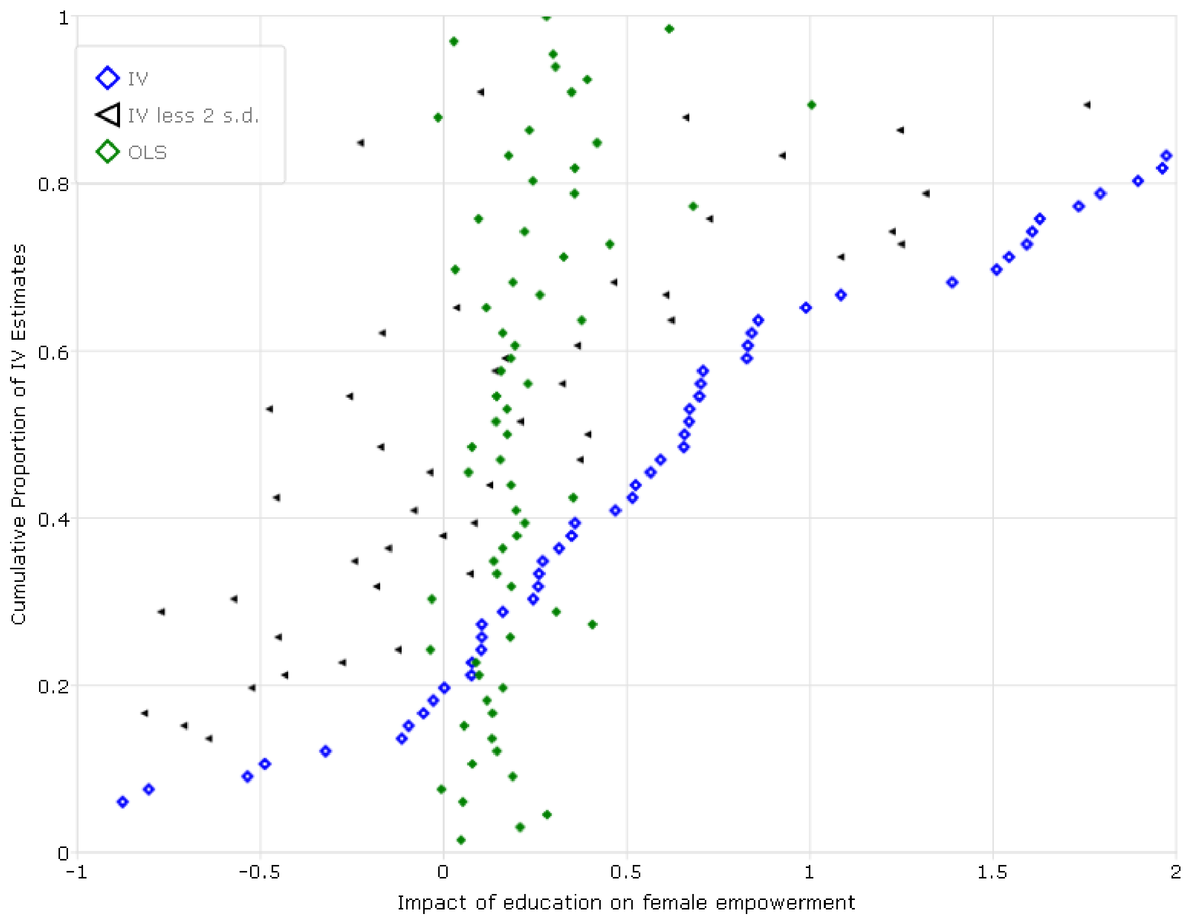
Note: EALOM IV estimates of basic education (the linear combination of the scaled coefficients on schooling and literacy) and respective standard errors from 128 DHS survey rounds in 54 different countries.

Figure G.5: Empirical cumulative distribution functions of the estimates of the impact of schooling (OLS) and basic education (OLS and IV) for women's empowerment



Note: For 128 country/year survey rounds from 54 countries the empirical cumulative distribution function for (a) OLS with "schooling alone", the scaled coefficient from bivariate regression of outcomes on years of schooling, (b)The OLS estimate of the "impact of basic education" which is the linear combination of the scaled coefficients of schooling and literacy, (c) the EALOM IV estimates of the "impact of basic education." Each set of estimates is sorted and hence each horizontal row (the nth estimate) is not from the same survey year.

Figure G.6: The empirical CDF of the EALOM IV and OLS estimates of the impact of basic education on women's empowerment, with IV 2 std error bounds



Note: (a) EALOM IV estimates of basic education (the linear combination of the scaled coefficients on schooling and literacy), (b) the IV estimates less two std errors, (c) the OLS estimates, all across 128 DHS survey rounds in 54 countries. Both the OLS and IV are sorted by the IV estimate so each horizontal "row" is a survey/year.

H Tables using individual women, regions, and EA to estimate schooling and learning

Table H.4: Estimating the association of schooling and learning with outcomes at the individual woman, regional, and EA level

	Fertility	Child Survival	Women's Empowerment
Woman level			
Schooling	-0.295	0.021	0.105
Reading	-0.114	0.008	0.097
Education (linear combination of primary schooling and reading)	-0.409	0.029	0.201
Obs. (women)	1055702	847655	386574
Country-year cells	129	129	129
Region level			
Schooling	-0.089	0.012	-0.171
Reading	-0.649	0.011	0.418
Education (linear combination of primary schooling and reading)	-0.738	0.023	0.248
Obs. (regions)	1471	1471	827
Country-year cells	129	129	129
EA level			
Schooling	-0.397	0.028	0.199
Reading	-0.226	0.008	0.168
Education (linear combination of primary schooling and reading)	-0.622	0.035	0.366
Obs. (EAs)	64975	64825	37519
Country-year cells	129	129	129

Note: The dependent variable is listed in the top row. Regressions are pooled at the level indicated in each panel.