World Development 122 (2019) 199–217

Contents lists available at ScienceDirect

World Development

journal homepage: www.elsevier.com/locate/worlddev

Educational gender gaps and economic growth: A systematic review and meta-regression analysis

Anna Minasyan [a,1], Juliane Zenker [b,1], Stephan Klasen [b,*], Sebastian Vollmer [b]

[a] University of Groningen, The Netherlands
[b] University of Goettingen, Germany

ARTICLE INFO

Article history:
Accepted 16 May 2019
Available online 10 June 2019

JEL Codes:
O47
I24
I25

Keywords:
Gender gaps
Education
Growth
Systematic review
Meta-analysis

ABSTRACT

Despite a large number of empirical studies, the controversy of whether a gender gap in education harms or boosts economic performance still persists. We conduct a systematic review and meta-analysis of the empirical literature on the link between gender inequality in education and per capita economic growth. After highlighting the methodological challenges of causally estimating this relationship, we document correlational evidence of a positive link between educational gender equality and economic growth. In particular, we find that studies that include male and female education as separate covariates in the growth regression report larger correlation sizes of female compared to male education with economic growth, except when an arguably problematic regression specification popularized by Barro and co-authors is used. Furthermore, studies that use gender gap (female/male ratio) in education as explanatory variable show an average partial correlation coefficient between growth and educational gender equality of 0.25, which is a moderate positive correlation. We also observe that the partial correlation increases with the use of initial education levels and social/institutional variables as controls, and becomes smaller with the use of country fixed effects, the inclusion of economic variables, and a higher share of female authors. We additionally assess six studies in our sample that use quasi-experimental methods (instrumental variable techniques) as an attempt to elicit a causal effect. Direction and magnitude of the estimates from these studies are comparable with the correlational evidence, but we note that the validity of many of the instruments used is open to question. We do not find evidence of publication bias in this literature.

1. Introduction

There are pervasive gender differences in different aspects of well-being and empowerment, including education, health, labor market opportunities, pay, political participation, and often also formal laws and informal social institutions (Klasen, 2016; Duflo, 2012). While some gender gaps are present in all countries of the world, they have been particularly sizable in developing countries, although some have been reduced substantially in recent years. Gender gaps in well-being and empowerment used to be seen primarily as issues of equity and justice. For example, the UN Convention on the Elimination of Discrimination of Women (CEDAW), concluded in 1977 and since ratified by nearly all countries of the world (although sometimes with reservations) is an example of this approach to the issue.

Starting in the 1990s, the relationship between gender inequality and development was beginning to be investigated. Initially, an important focus was on the strong empirical link between female education and fertility as well as child mortality (e.g. Summers, 1994, Murthi, Guio, & Dreze, 1995). Soon thereafter first studies appeared that investigated the relationship between gender gaps and economic performance (e.g. Hill & King, 1995). An increasing number of studies then started to rely on cross-country growth regressions that had been pioneered in the early nineties (Barro, 1991). While there are studies that examine the link of gender gaps in employment, pay, health, laws, and empowerment on economic growth, the largest number of studies to date has focused on the relationship of gender differences in education and economic growth. This is related to human capital being a key ingredient of growth theory and growth empirics so that education always features prominently in such growth analyses, and it is not a big leap to disaggregate education by gender. Moreover, there are widely

* Corresponding author.
E-mail addresses: a.minasyan@rug.nl (A. Minasyan), jzenker@uni-goettingen.de (J. Zenker), sklasen@uni-goettingen.de (S. Klasen), svollme@uni-goettingen.de (S. Vollmer).

[1] First authors.

https://doi.org/10.1016/j.worlddev.2019.05.006
0305-750X/

available and quite reliable metrics of education quantity by gender, including enrolment rates, literacy rates, and years of schooling (e.g. Barro & Lee, 2013).

On the empirical side, in the early nineties, Barro and Lee (1994) and Barro and Sala-i-Martin (1995) reported the 'puzzling' finding that more female years of schooling is associated with lower economic growth, while the reverse was the case for male schooling. Many other studies, however, found the opposite and several studies were published to explain how the unexpected findings from Barro and co-authors had come about (e.g. Dollar & Gatti, 1999; Klasen, 2002; Lorgelly & Owen, 1999, and Knowles, Lorgelly, & Owen, 2002).

Despite the large number of empirical studies that examined this topic, the controversy whether the gender gap in education harms or boosts economic growth still persists. This review and meta-analysis study aims to systematically assess the evidence and synthesize the differing and partly opposing findings (Stanley & Jarrell, 1989; Stanley, 2001), while also discussing the challenges of causally estimating the impact of gender gaps on economic performance.

The body of the relevant literature we identify consists of cross-country studies (including pure cross-sections or panels, 34 studies), single country time series studies (17 studies), and single country cross-regional studies (3 studies). We group these studies accordingly for our analysis. The cross-country studies are further divided into *comparative* (17 studies, 168 estimates) and *gap* (17 studies, 216 estimates) studies. For compatibility reasons we conduct meta-regression analysis for both sets of studies separately. We use weighted OLS (clustered at the study level) as well as a Random Effects Maximum Likelihood estimator to study the average partial correlation of the educational gender gap with growth.[2]

Our results show that in the *comparative* studies, which use male and female education as separate covariates in the growth regression, show a larger positive association of female compared to male education on growth, except when a regression specification popularized by Barro and co-authors (e.g. Barro & Lee, 1994) is used. We consider this specification to be problematic as it is likely to assign unrelated region-specific growth factors to gender inequality in education. For the *gap*-studies, which use the female-male education ratio or difference in the growth regression, we find evidence for positive and statistically significant relationship between gender equality in education and growth. We document an average partial correlation coefficient of economic growth with the ratio of female over male education of 0.25, which is a moderate size. The average partial correlation does not appear to be influenced by publication bias. Further, it increases when one controls for initial education levels and social/institutional controls, while it falls with the use of country fixed effects, the inclusion of economic controls, and the share of female authors. Evidence from the single country cross-regional studies, by and large, confirms the positive association of educational gender equality with growth.

Given the identification challenges in this type of literature, we also do a risk of bias assessment of six studies that use instrumental variable (IV) methods and attempt to lend some measure of support to a causal interpretation of this relationship. Direction and magnitude of the estimates from these studies are comparable with the correlational evidence. However, our assessment of the merits of the instruments used in these six studies reveals that the validity of many of the instruments is open to question.

Finally, the time series analyses that we investigate are based on a few single countries and generally use much weaker methods. We, therefore, refrain from drawing any generalized conclusions from this set of studies. Just to note, the literature from the time series studies also suggests that reducing gender inequality in education is associated with larger economic growth. In the next section we discuss the conceptualization of the relationship between gender equality in education and economic growth.

2. Conceptual framework

The existing theoretical arguments highlight both mechanisms for a positive as well as a negative effect of educational gender gaps on economic growth, which implies that it remains an empirical question whether the negative effects outweigh the positive ones and if this is universally true or context dependent. There are two arguments that suggest that gender gaps in education could actually promote economic performance. On the one hand, Becker (1981) essentially argues that there are (static) efficiency gains to a sexual division of labor where each gender specializes on the tasks where they have a comparative advantage, which Becker sees for women in home production (due to the complementarity of child-bearing and child-rearing). A second argument can be made following suggestions by Tertilt and Doepke (2014): due to different gender roles, higher female education (and associated higher employment and earnings) could lead to more household consumption rather than savings which could serve to lower economic growth.[3]

On the other hand, there is a substantial number of papers arguing the reverse, i.e. that gender gaps in education can reduce economic performance. First, this theoretical literature suggests that such gender inequality reduces the average amount of human capital in a society and thus harms economic performance. It does so by artificially restricting the pool of talent from which one can draw for education and thereby excluding highly talented girls (and taking less talented boys instead, e.g. Dollar & Gatti, 1999; Cuberes & Teignier, 2016). Moreover, if there are declining marginal returns to education, restricting the education of girls to lower levels while taking the education of boys to higher levels means that the marginal return to educating girls is higher than that of boys, and reducing the gap would boost overall economic performance. Such an effect would be exacerbated if males and females are imperfect substitutes (World Bank, 2001; Knowles et al., 2002).

A second argument relates to externalities of female education. Promoting female education is known to reduce fertility levels, reduce child mortality levels, and promote the education of the next generation. Each factor in turn can have a positive impact on economic growth (World Bank, 2001; King, Klasen, & Porter, 2009). Some models emphasize that there is a potential of vicious cycles with larger gender gaps in education or pay leading to high fertility, which results in poverty among the next generation leading to low-income poverty traps (e.g. Galor & Weil, 1996; Lagerlöf, 2003). But there is also an important timing issue involved here. Reducing gender gaps in education can lead to reduced fertility levels which will, after some twenty years, lead to a favorable demographic constellation which Bloom and Williamson (1998) refer to as a 'demographic gift'. For a period of several decades, the working age population will grow much faster than overall population, thus lowering dependency rates with positive repercussions for per capita economic growth.[4]

[2] Although common to the meta-analysis terminology, we do not make use of the terms *effect* and *effects sizes* in our empirical analysis in order to acknowledge that our research synthesis is based on observational studies that mostly do not allow for a causal interpretation.

[3] Tertilt and Doepke (2014) relate this argument mainly to gender-gaps in earnings and unearned incomes.

[4] See Bloom and Williamson (1998) and Klasen (2002) for a full exposition of these arguments.

A third argument relates to international competitiveness. Many East Asian countries have been able to be competitive on world markets through the use of female-intensive export-oriented manufacturing industries (e.g. Seguino, 2000a, 2000b).[5] In order for such competitive export industries to emerge and grow, women need to be educated and there must no barrier to their employment in such sectors. Gender inequality in education and employment would reduce the ability of countries to capitalize on these opportunities (World Bank, 2001; Busse & Spielmann, 2006).

Hence, given these competing arguments, it becomes an empirical question whether and to what extent gender inequality can have an impact on economic growth. As the different models suggest different mechanisms, ideally one would look into these mechanisms in the empirical literature. Our meta-regression can partly address this question by examining the role of particular control variables – some of which represent mechanisms.

3. Empirical challenges

Linking gender inequality in education to aggregate economic output empirically poses a range of challenges. Difficulties concern a number of measurement as well as methodological choices and identification strategies, some of which we briefly discuss below.

First, while one can measure the effects of female education or educational inequality on income at the micro level, it is unclear that they carry over to the macro level. One can readily estimate how household incomes increase if individual women have higher education using Mincer-type wage regressions. Whether this effect is the same if *all* women increase their education, is less clear for several reasons. On the one hand, it might be the case that the macro returns to increasing female education are much smaller than the micro returns calculated in wage regressions. For example, to the extent that education is merely a signal for higher ability rather than representing greater human capital, it would not increase output. Similarly, the higher education might boost female earnings but lead to no growth if they work in sectors that engage in lucrative, but for the aggregate economy unproductive activities (such as adding to a public sector that increases hold-up problems for firms and individuals, Pritchett, 2001).

On the other hand, a range of positive spillovers have been suggested that are difficult to capture at the micro level. In particular, there are the well-known associations of female education with lower fertility and lower child mortality, which can boost economic performance through demographic channels as well as improved human capital (Klasen, 2002; King, Klasen, & Porter, 2009). It is conceivable, that efforts to boost female education can even lead to a wider social change that could, for example, reduce cultural and social barriers to female economic activity and further boost economic performance. This also suggests that measuring the impact of policies to reduce gender inequality at a micro level using, for example, Randomized Controlled Trials (RCTs) can at best capture some local spillovers. RCTs might therefore underestimate effects that would be expected if such programs were implemented to scale. This is particularly the case if such effects materialize with some delay. In short, while micro level studies provide important insights into mechanisms and local effects, they may not be a reliable guide to aggregate economic effects. Here aggregate analyses, such as cross-country or cross-regional regressions, have distinct advantages, although they suffer from other problems which we discuss below.[6]

Second, when measuring the relationship of gender gaps and economic performance, one has to recognize the further difficulty that a considerable share of women's work is not captured by national accounts and included in GDP or GNI estimates. As is well known, unmarketed services such as care for own children, elderly, or housework which is done overwhelmingly by women is not included in the System of National Accounts (SNA) and would make up a significant share of GNI if it was included (UN, 2008; UNDP, 1995; OECD, 1995; McKinsey, 2015). While the omission itself is a problem, the interaction with gender gaps poses a particular challenge. It may be the case that countries with higher female education and associated higher employment rates have a higher GDP partly because in those countries more of the household service and care work is marketed and thus included in GDP, while the unmeasured service work done by women in their own households is correspondingly smaller (UNDP, 1995). If this is the case, a coefficient of female education on GDP might overestimate the impact education equality has on the *total* provision of (marketed and unmarketed) services, an important indicator of welfare. We know of no work that has empirically attempted to assess this potential bias.

Third, while we have noted some distinct advantages of using aggregate-level analyses such as cross-country regressions, there are other problems that can bias estimates and preclude a causal interpretation. Endogeneity is the most serious problem. In particular, when examining the relationship between gender inequality in education and economic performance, reverse causality can be an issue and unobserved heterogeneity is bound to be a problem (Bandiera & Natraj, 2012; Dollar & Gatti, 1999). Choosing an appropriate time structure for dependent and independent variables, done in most of the studies reviewed below, can partly address reverse causality, and some time series studies investigate Granger causality, another partial remedy. Using country fixed effects in a panel setting can deal with time-invariant unobserved heterogeneity. While the lags and high saturation with fixed effects can lend some support to the interpretation of the coefficient as a causal relationship, there remain possibilities of confounding by time-varying but inter-period persistent unobserved confounders. Adding covariates that might proxy for unobserved heterogeneity can further reduce the problem but runs against multicollinearity and/or degree of freedom issues and may not capture the most important left-out variable. Some panel studies use difference or system GMM estimators to address endogeneity, but these approaches are often rather unstable and rely on assumptions that are hard to test (e.g. Roodman, 2009). Other papers use IV strategies to address endogeneity, but the instruments used are controversially discussed (e.g. Bandiera & Natraj, 2012). Given the debates about credible identification of causal effects in cross-country regressions, relying only on studies with a broadly accepted credible identification strategy, as done in some other systematic reviews (e.g. Zürcher, 2017) would essentially lead to an empty set. Instead, to approximate the severity of the problem, we make an attempt to compare findings from our main analysis with those from studies that use more rigorous analytical approaches, i.e. quasi-experimental – precisely, instrumental variable (IV) – set-ups. However, an assessment of the validity of the instruments used in the latter studies, still casts some doubt that

[5] Klasen (2006) reviews the literature and also notes that such strategies have now been extended, with some success to countries such as Tunisia, Bangladesh, China, and Vietnam.

[6] Simulation results based on calibrated theoretical models, which tend to explicitly consider spill-overs and general equilibrium effects can also address this problem (e.g. Cavalcanti and Tavares, 2016; Cuberes and Teignier, 2016). But, of course, theoretical models (and their derived simulated results) tend to focus on few mechanisms and thus are also limited by their set-up to those which may or may not be empirically relevant.

all remaining potential biases are removed to allow for a causal interpretation of the findings.

Thus, this systematic review and meta-analysis cannot provide a definitive assessment of the causal impact of gender inequality in education on economic growth but rather they reveal the evidence in line with the methods used in the existing literature.

4. Systematic-review methodology

4.1. Criteria for the inclusion of studies

We follow Petticrew and Roberts (2006) and use the PICOS model (Population, Intervention, Comparison, Outcome and Setting) to define the inclusion criteria for our review as detailed below.

Population. We include all quantitative cross-country and within-country cross-regional studies that relate the educational differences between males and females in the whole population (based on survey or census data) to an indicator of economic performance. We include studies published in refereed journals as well as discussion papers and control for the publication status in our review.

Intervention. We include studies that use changes and/or levels of gendered educational variables as an explanatory variable in regression analysis based on observational and macroeconomic data. That is, on the right-hand side of the estimation equation should be either the levels or changes of female and male education separately (both in one regression) or a measure of the gender gap in education. We include all measurements of educational quantity (enrolment, attainment, years of schooling) as well as studies that include instrumental variables for the gendered educational variable and time lags of these.

Comparison. We compare only quantitative, observational studies that include regression analyses on the relationship between a gendered educational variable on the *outcome* specified below. The studies should have a clearly defined sample, method, and results description. Comparison is based on educational gender differences between countries as well as changes within a country over time. Based on the research design, we categorize the studies into the following groups:

a) Within-country time series studies: These studies use time series econometric techniques to relate a time series of gendered educational variables to a time series of economic growth in a particular country. While these studies are summarized in the systematic review, we do not include them in the meta-analysis.
b) Cross-country and cross-sectional regression studies: These studies use variation between countries.
c) Panel cross-country studies: These studies use variation across countries and over time.
d) Cross-regional studies. In the systematic review (but not in the meta-analysis) we also include the few available cross-regional studies that exploit variation between regions within a country (and sometimes also over time).

Outcome. We solely include studies that try to explain economic growth. Thus, we include studies with the outcome variable defined by the growth rate of GDP per capita even if in some cases the outcome is the level of per capita income measures, given the study design allows to translate this to economic growth (i.e., country fixed effects). We exclude (the very few) studies that only consider aggregate income or economic growth (instead of per capita income or per capita growth) and do not at the same time control for population (or population growth).

Setting. We focus on aggregate-level outcomes (at the country or region level). The studies must include a regression analysis.

4.2. Search strategy

In order to make the search and inclusion of the literature as transparent as possible, we use easily accessible, disciplinary as well as cross-disciplinary general research databases as well as reference snowballing techniques (backward and forward citation) to collect literature on the relationship of gender inequality in education on the economic growth. In particular, reference snowballing is recommended by Petticrew and Roberts (2006) as well as Waddington et al. (2012) for overcoming the problems in searching social science literature.[7]

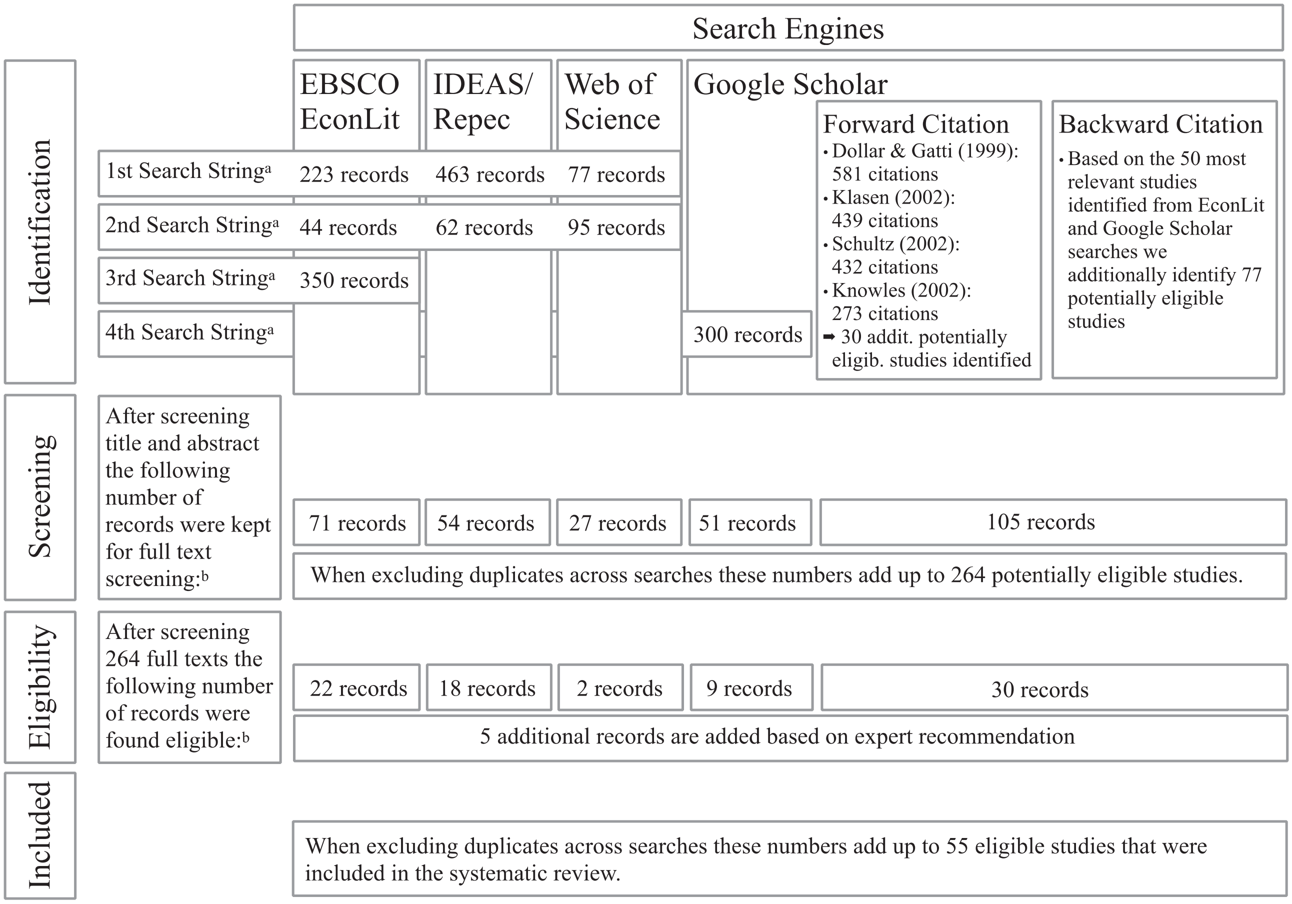
In particular, we used four easily accessible research databases: EconLit, IDEAS, Web of Science and Google Scholar. The first two contain articles from the discipline of economics, while the latter two include all disciplines. EconLit includes close to the universe of published articles in economics journals (including many relatively unknown journals), in addition to selected highly reputable working paper series (such as the NBER series). IDEAS is the largest bibliographic database for studies in Economics and, complementary to EconLit, also covers grey literature (e.g. a large number of departmental working paper series, etc.). Web of Science, additionally, covers published research articles across all social science disciplines. All three databases allow for sophisticated Boolean-phrased search strategies in titles, abstracts, and full texts. Furthermore, we use Google Scholar, which applies an entirely different search concept. While the search engine only allows for a simple combination of search terms, it provides a relevance ranking based on a complex set of built-in sorting criteria. Furthermore, Google Scholar allows for tracking citations in forward and backward directions and allows for full text searches, which we made use of. As Google Scholar usually generates thousands of references (and presenting them in declining order of relevance), we limited ourselves to the most relevant studies identified (see below). Our search strategy is structured based on the main concepts examined in the review, which are education, gender (-gap, male, female, and etc.), and economic growth. We combine three to four sets of synonymous terms in several ways to capture all potentially relevant studies. See the Supplementary Material for a detailed overview on all applied search strings and search specific results.[8] As Boolean-phrasing is not possible in Google Scholar, the search was carried out for a simple combination of following keywords in the text, details are also provided in the Supplementary Material.

As detailed in Fig. 1 below, the search strings in EconLit yielded a total of 617 papers (many of which were duplicates), in IDEAS we found a total of 525 records, in web of science 172. The search in Google Scholar resulted in 26.500 studies, which mention all of the keywords in the text. Relevance declined sharply after the first 300 articles. No restriction on time/year and language were put on any of the above searches and we retained the 300 first studies.[9]

[7] For example, estimation method filters or keywords do not necessarily appear in title or abstracts of papers in economics while it is quite straightforward and expected in the health literature.

[8] To increase the chance of capturing all relevant studies, we used two different search strategies in the databases. One used a combination of search terms that had been found through experimentation to yield a particularly high share of (potentially) relevant studies: (education* *equality gender* growth*) OR (education* gap* gender* growth*) OR (education* female* growth*) OR (school* female* economic growth*) OR (school* girl* economic growth). The other was built up systematically from all combination of the three or four parts of our search (a synonym each for education, gender, and growth, complemented by *equality). It turned out that both strategies eventually converged to a very similar set of eligible studies. See Supplementary Material for details.

[9] But our English search terms will implicitly focus on English-language studies except when non-English studies include English abstracts, title, and keywords. In the end, all included studies are in English.



[a] For detailed motivations of the four different search strings see text.
[b] Double blind screening by two reviewers was done based on previously defined criteria described in 1a) to d) in the text. Figures presented in the first row are net of duplicates within search engine but not across search strings.

Fig. 1. Overview of the literature search.

Additionally, we examined the reference lists of 50 particularly relevant and recently published articles, adding all (77) therein cited additional studies (i.e. not previously identified studies) to our literature database. Further, forward citation was carried out for the most cited papers as of January 28, 2016 in Google Scholar in gender inequality in education and growth, which are:

- Dollar and Gatti, World Bank Working Paper 1999 (581 citations) – 17 new ones added
- Klasen, World Bank Economic Review 2002 (439 citations) – 6 new ones added
- Schultz, World Development 2002 (432 citations) – 1 new one added
- Knowles, Lorgelly and Owen 2002 Oxford Economic Papers (273 citations) – 6 new ones added

In this step, using Google Scholar citation tracking, all references have been reviewed in which the aforementioned studies have been cited. In total, 30 additional papers were added to the collection through this procedure. In sum, all searches resulted in a total of 1421 potentially relevant records, which were then passed along for screening. Screening was done in two steps, based on the criteria described in section 4.1 by two reviewers independently.

In the first screening, titles and abstracts were screened, with an aim to remove those records which were clearly not relevant for the review based on the criteria above in section 4.1. This led to 308 relevant studies. The removal of duplicates across searches led to a reduction to 264 studies. For the remaining 264 studies, we carried out a full-text screening, completed independently for each study by two reviewers. Thereafter, the bibliographic data was extracted and 264 studies were assessed by the two reviewers independently whether the study reported original regression results (Yes = 1, No = 0), whether the study reported a regression that had per capita income or income growth as a left-hand side variable (Yes = 1, No = 0), and whether in the same regression right-hand side variable(s) were included representing a gap in education or education measures disaggregated by gender (Yes = 1, No = 0). If any of these criteria was coded with zero the study was rated as irrelevant for our review, otherwise it was rated as relevant. Any cases classified as unclear were discussed together with a third reviewer (an expert on the topic) for a final inclusion decision.

In the last steps we combined the two reviewer's eligibility assessments and discussed the unclear cases among the entire team, and added five additional records based on expert recommendations. In result, 55 studies published in journals, as working papers, as books, or doctoral theses were selected as relevant for the synthesis and a large amount of papers was excluded due to one or more of the following reasons: they were solely theoretical; had only descriptive results (means and/or scatterplots); did not have per capita economic growth or level of income as the dependent variable; did not have a gap/ratio of male and female education as the explanatory variable; did not have female and male education as the explanatory variable. The search history has been documented on user accounts and the excluded studies with abstracts and data can be retrieved when necessary.

Out of the 55 studies eligible for this systematic review, 39 are published journal articles, 13 are working papers, one is a book chapter, one is a conference proceeding, and one is a dissertation. All studies are listed in a separate reference section below. Seventeen of the studies use time series methods for single countries as in (a) in section 4.1, one study uses Bayesian model averaging, three studies run within-country cross-sectional regressions as in (d) in section 4.1, while the remaining 34 studies cover a larger set of countries using cross-section or panel methods as in (b) or (c) in section 4.1. For comparability reasons, we include only these 34 cross-country studies for the meta-analysis presented in sections 6 and 7. The time series studies are summarized in Appendix 1 and in Appendix 2, we summarize the three studies that run sub-national regressions using male and female education as covariates separately, and one Bayesian Model Averaging Study that also uses disaggregated education measures.[10]

5. Meta-regression analysis methodology

5.1. Data extraction and sample description

The 34 studies that are identified as eligible and included in the meta-analysis report a total of 383 regression equations that investigate the relationship between educational gender equality and growth. Data extraction for all studies was done on the coefficient level of individual regressions, as many studies do not just report one estimate but contain multiple coefficients of different regressions that are relevant for our assessment. For each relevant regression we extracted information on coefficient-related characteristics (e.g. standard error, t-statistic, p-value), dependent variable, explanatory variables, data type, source and period, and estimation method. For a detailed overview of the extracted criteria see the Supplementary Material.

The question whether a gender gap in education is associated with economic growth is assessed in two common ways in our sample. As shown in Table 1, half of the studies, and 168 estimates, are based on gender-disaggregated measures for education (i.e. one measuring a country's male and one measuring a country's female education), which are included separately in the analysis. For simplicity, we will refer to these as *comparative* studies. The other half of studies, or 216 estimates in our sample, are based on regression equations that use the disaggregated measures to create a, gender gap", i.e. they combine the two disaggregated measures to a single variable by constructing a difference or ratio between the two, and eventually include the resulting gap-variable in the analysis (female over male ratio).[11] We will refer to these as the *gap*-studies. As these approaches are fundamentally different and the estimates are not comparable, we perform separate analyses for each set of studies, respectively.[12]

[10] There, the evidence shows that reduction of gender inequality in education has, overall, a positive correlation with economic growth. We do not include the time series studies in the meta-analysis because of comparability issues, high heterogeneity in methods, with dubious methods of high risk of bias. We also are concerned about external validity as all of these time series studies focus only on few countries, with more than half the studies being about Pakistan.

[11] One of the studies presents regression analysis for both, the gap and the disaggregated measures (Knowles et al., 2002).

[12] Transforming the coefficients from the studies using disaggregated indicators into female-to-male education ratios in order to include all studies in one meta-analysis would require sufficient information about the variance-covariance relationships of the two regressors. As we do not have this information for most studies we refrain from such an exercise.

Table 1
Methods used in the studies included for meta-analysis.

Data	Method	# of studies	% of studies
Cross-section	OLS	13	0.38
	IV	5	0.15
Total cross-section		18	0.53
Panel	Pooled OLS	8	0.24
	RE, FE, SUR	13	0.38
	IV, GMM	15	0.44
	Other	8	0.24
Total panel		23	0.68
Total		34	1

Notes: Please note that adding the studies using different methods, as well as adding the total number of cross-section and panel studies leads to numbers that exceed the total number of studies. This is due to the fact that some studies use cross-section as well as panel data and many papers use several methods in different sets of regressions.

Most studies in our sample report coefficients from more than one method: 13 studies report results from cross-section ordinary least squares (OLS) regressions, five report results from cross-section instrumental variable (IV) regressions, eight report results from pooled OLS panel regressions, 13 report results using random effects (RE), fixed effects (FE), or seemingly unrelated regression (SUR) panel methods, 15 report results from panel IV regression or using generalized methods of moments (GMM), see Table 1. Eight studies report coefficients from other panel regression methods, which do not clearly fall into the former categories, i.e. Extreme Bound Analysis, Bayesian Averaging of Classical Estimates, Three Stage Least Squares, Chamberlaińs Pi-matrix, Iteratively Reweighted Least Squares, and Semi-parametric Partially Linear Regression.

5.2. Summarizing regression coefficients

In order to make regression coefficients comparable across regression equations and studies, we convert the extracted beta coefficients into partial correlation coefficients – a measure that indicates to which extent two variables are associated and which direction this association takes, while holding other variables constant (Stanley & Doucouliagos, 2012). We calculate the partial correlation coefficient r as

$$r_{ij} = \frac{t}{\sqrt{t^2 + df}} \quad (1)$$

based on regression i in study j. Further, t denotes the t-statistic of the relevant regression coefficient (i.e., the gender gap) and df denotes the degrees of freedom in each regression. The standard error of the partial correlation coefficient is calculated as $SE_r = r/t$. The partial correlation coefficient is a standardized statistic of correlation – it is scale-less, which enables us to easily compare coefficient sizes across multiple studies and regressions.

We rely on two established methods to run the meta-analysis by pooling the obtained partial correlation coefficients in order to identify the true size and sign of the underlying correlation. These methods are *fixed effects* and *random effects* meta-regression analysis (MRA) suggested by Brockwell and Gordon (2001) and Stanley and Doucouliagos (2015, 2016) for studies in the economics and business disciplines. The *fixed effects* model assumes that any existing difference in the partial correlation coefficients across studies are due to idiosyncratic study-level errors (Borenstein et al. 2010), or that studies can be considered as homogenous. The left-hand side variable in the model is then the partial correlation coefficient, while the right-hand side comprises the true underlying average correlation (i.e. a constant) as well as an error term:

$$r_{ij} = \beta_0 + e_{ij} \quad (2)$$

This equation can be further augmented with weights that reflect precision in the estimates. Hedges and Olkin (1985) suggest the most optimal weight to be the inverse variance, $w_i = 1/SE_i^2$, where SE_i^2 is the square of the standard error of each estimate in the sample (see also Stanley and Doucouliagos 2012). While the fixed effects model is the most intuitive form of synthesizing research findings in our sample, it does not account for the fact that observational macroeconomic studies greatly differ in terms of sample composition, estimation method, periods, and specification. It is likely that the true underlying correlation size varies with these study characteristics. We, therefore, augment our model in (2) by estimating *random effects* MRA – which relaxes the assumption that all the estimates in our sample are drawn from only one population with the same mean. In other words, in addition to within-study errors, we also allow for errors generated from between-study differences and heterogeneity between studies.

Namely, we use the Random Effects Maximum Likelihood (REML) estimator, which controls for the between-study variance (Thompson & Sharp, 1999, Benos & Zotou, 2014, Gallet & Doucouliagos, 2017).[13] The weights in this case can be expressed as $w_i = 1/SE_i+\tau^2$, where τ^2 is the between study variance (Thompson & Sharp, 1999, Borenstein et al., 2010, Stanley & Doucouliagos, 2015, 2016). As we use multiple estimates from the same study and it is possible that within-study errors are not independently distributed (i.i.d), we further cluster the standard errors at the study-level.

5.3. Publication bias

One key purpose of meta-regression analysis (MRA) is to detect publication bias in the relevant body of literature. Publication bias may arise from several sources, like predispositions or expectations regarding certain test results on the side of the authors, reviewers, or the editor (Stanley & Doucouliagos, 2012). Moreover, studies that find statistically significant results (which implies relatively smaller standard errors) are more likely to be published (Stanley, 2005). MRA identifies the existence of publication bias in the literature by pooling all estimates and examining the distribution of these estimates graphically (*funnel plot*) and by formally testing for *funnel asymmetry* (Stanley, 2005, Duval and Tweedie, 2000, Egger et al., 1997).

Following Stanley and Doucouliagos (2012), we specify the test for funnel asymmetry in a regression as follows:

$$r_{ij} = \beta_0 + \beta_{se}SE_{ij} + e_{ij}, \quad (3)$$

where r_{ij} is again the partial correlation coefficient and the constant term β_0 again represents the genuine average correlation between gender education gap and economic growth. *SE* is the standard error of the partial correlation coefficient, while e_{ij} is the error term clustered at the study level. Based on this equation we employ the FAT-PET test, which comprises of two jointly tested hypotheses. First, $H0_{FAT}$: $\beta_{se} = 0$, formally tests for funnel asymmetry (FAT) in Fig. 4, i.e. publication bias. The rejection of $H0_{FAT}$ is an evidence for biased reporting of results by giving preference to those with statistical significance. Moreover, $H0_{PET}$: $\beta_0 = 0$ tests for the existence of a statistically significant average correlation conditionally on controlling for a possible publication selection, or the precision-effect test (PET). However, Stanley (2008) reports that β_0 in Eq. (3) may be biased downward when $H0_{PET}$ is rejected. To overcome this problem, we follow the recommendation of Stanley and Doucouliagos (2012) and further use a non-linear estimator by replacing the standard error, *SE*, with its square term, SE^2. In this case β_0 is called the precision effect estimate with standard error (PEESE) based on the following equation:

$$r_{ij} = \beta_0 + \beta_{se}SE_{ij}^2 + e_{ij} \quad (4)$$

5.4. Heterogeneity

Furthermore, we augment Eq. (3) with a vector of moderator variables to explain the heterogeneity in the coefficient sizes, r_{ij}. We presume that the true underlying correlation size varies with characteristics regarding model specification (e.g included covariates), regression method (e.g. OLS cross-section regression, fixed effects panel regression, instrumental variable methods) and differences in measurement of the education variable. We extend Eq. (3) and estimate it as follows

$$r_{ij} = \beta_0 + \beta_{se}SE_{ij} + \sum \alpha_k Z'_{ij} + e_{ij}, \quad (5)$$

where *Z* is the set of moderator variables that includes the relevant study and regression characteristics. Details on the variables included are discussed in section 7.

6. Comparing female and male education coefficients (“*Comparative* studies)

Seventeen studies in our sample run growth regressions by separately including the female and male education as explanatory variables on the right-hand side. Due to the regression structure of these studies, the relationship between educational gender inequality and economic growth cannot be investigated directly as the information on the variance-covariance matrix is not available for each regression. Yet, inference can be made by comparing the two sets of coefficients descriptively and graphically.

6.1. Descriptive evidence and the Barro-‘Effect’

In our sample, 168 regressions include female and male education variables separately in one regression. In 20 percent of these ‘comparative’ regressions the female education coefficient is positive and statistically significant and in 14 percent of the regressions it is larger than the male education coefficient and statistically significant at the conventional level, see Table 2. From these purely descriptive results, one might conclude that only a minority of studies suggest that female education is positively associated with economic growth, and even a smaller minority that it does so more than male education. In fact, male education is positive and significantly related to economic growth in more than three times as many studies. In 48 percent of regressions the relationship of female education with economic growth is significantly negative.

The puzzling result, that female education is seemingly negatively correlated with economic growth, while the correlation with male education is positive and statistically significant, was first reported in an influential study by Barro and Lee (1994). This study and later studies following this approach were criticized by the follow-up research for three distinct features: in Barro’s specification, used by him and his co-authors, and others in several papers (see below), the regressions did not control for time-invariant characteristics at the country or regional level (using dummy variables or fixed effects) when using panel data; did not control for regional specificities (using regional dummy variables) when using cross-sectional data; and the education variables were included as the base value of the usually averaged growth periods instead of

[13] Stanley and Doucouliagos (2015, 2016) argue that unrestricted fixed effects WLS performs better in the presence of publication bias and, in the absence of this bias, the unrestricted fixed effects WLS performs as good as random effects.

Table 2
Comparative studies –Descriptive summary of results.

	Coefficient			
Indicator	Positive, significant	Negative, significant	>Male coefficient, significant	>Female coefficient, significant
Male education	0.70	0.10	–	0.64
Female education	0.20	0.48	0.14	–

Notes: Total number of estimates is 168. *Significant* refers to statistical significance at least at the 10 percent level (p-value < 0.1).

the period average. These features appear to drive the conclusions on the negative relationship of female education with growth.[14]

More precisely, several authors suggest that the negative association of female education on economic growth in these Barro-style regressions may be an artifact of certain regional experiences that lead to omitted variable bias. Dollar and Gatti (1999) emphasize that levels of female education were relatively high in Latin America already at the beginning of the study period of most regressions (usually 1960–1970). At the same time, per capita growth was low over the study period, especially if it included the crises periods of the late 1970s, 1980s, and early 1990s. They suggest including a Latin American dummy to the regression to overcome this omitted variable bias. Lorgelly and Owen (1999) further document that high initial gender gaps in certain fast-growing East Asian economies contribute to the "puzzling" result in a similar vein. To overcome this problem Knowles et al. (2002) suggest the use of education period averages instead of base values and show that this leads to a reversed relationship of the educational variables with growth. Alternatively, the use of regional dummy variables for Latin America and East Asia could (partly) overcome this omitted variable bias, or both dummy variables and period averages can be used. Taken together, this would imply that using initial year education data and failing to control for regional (or country) dummy variables would relate the low growth in Latin America to high initial female education there, and conversely high growth in East Asia to comparatively low initial female education. Clearly, this is a dubious attribution as many factors other than initial female education contributed to the East Asian economic miracle (e.g. World Bank, 1994) and Latin America's poor growth record (e.g. Taylor, 1998), and this is thus most likely a problem of bias due to unobserved heterogeneity.[15]

In our sample a number of studies replicate the Barro specification, i.e. also use initial educational values, do not control for time-invariant country heterogeneity, and do not include regional dummies. In Table 3 we list the number of estimate pairs obtained from Barro-type regressions versus those that deviate from it, for instance, by including education as period average, controlling for time invariant characteristics with fixed effects or in a GMM set-up, or including regional dummies in the regression, all of which thus attempt to tackle the problem of unobserved heterogeneity at the country or regional level.

6.2. Graphical analysis

We plot coefficient relations that originate from typical Barro-style versus those that originate from other regressions in the sampled studies to understand whether our descriptive results in Table 2 are in fact driven by the typical Barro-style specification. We calculate partial correlations of each of the two education coefficients with the growth variable (as described in Eq. (1) above) and plot the relationship of the resulting coefficient pairs; see Figs. 2 and 3, which show the full set of estimates and the within-study averages of estimates, respectively. Similar to the descriptive results in Table 2, we find a large cluster of coefficients in the left upper corner in both figures, suggesting that male education is positively associated with growth while female education it associated negatively. Yet, when investigating the studies more closely, it becomes apparent that Barro-style specifications (green dots) drive the vast majority of coefficient pairs in the upper left quadrant, replicating Barro's 'puzzling' result.[16] Fig. 3 also shows that study-average estimates using Barro-style specifications drive most results in the upper left quadrant.

The green dashed lines in Fig. 3 additionally represent precision effect estimates of (squared) standard errors (PEESE) of the male and female coefficients for the Barro-style regressions and non-Barro style regressions, respectively. It becomes evident that the Barro-specifications dominate the plots as they are associated with male-positive (PEESE: 0.154; p-value <0.01) versus female-

Table 3
Number of Barro-style specification per study.

Study	Non-Barro	Barro	Total
Barro and Lee (1994)	1	20	21
Barro (1996a)	1	4	5
Barro (1996b)	0	4	4
Caselli et al. (1996)	4	2	6
Cooray et al. (2014)	16	0	16
Cooray and Mallick (2011)	21	0	21
Dollar and Gatti (1999)	2	0	2
El Alaoui (2015)	6	0	6
Forbes (2000)	6	6	12
Hassan and Cooray (2015)	6	0	6
Huffman and Orazem (2004)	1	0	1
Kalaitzidakis et al. (2001)	5	0	5
Knowles et al. (2002)	18	0	18
Lorgelly and Owen (1999)	0	6	6
Perotti (1996)	1	14	15
Seguino (2000)	4	0	4
Szulga (2006)	7	13	20
Total	99	69	168

[14] In one regression, he includes regional dummies and then the negative correlation disappears.

[15] Further, Klasen (2002) notes that estimating the gap-growth relationship might be further complicated by multicollinearity issues. He emphasizes that the two education variables are highly correlated in most countries (with correlation coefficients usually exceeding 0.9) and that large standard errors of estimated coefficients as well as the sudden reversal of the coefficient signs in different specifications manifest the possibility of a multicollinearity bias. This is addressed below in the studies using ratios of male and female education.

[16] As can be seen in Fig. 2, there are a few Barro-style regressions showing a positive correlation of female education with growth. In his initial study, Barro and Lee (1994) reports – but not further discusses – that the relationship of growth and female education turns positive once logged fertility and population growth are included as control variables. A possible explanation for this finding may be single influential observations with high GDP growth, population growth, and fertility rates but low initial female education. For instance, Botswana experienced exceptional growth rates in GDP over the study period as well as high initial fertility and population growth rates on the one hand, while starting off with extremely low levels of female education on the other. If the negative relationship between initial female education and economic growth is driven by this outlier it would be conceivable that the fertility and population growth variables pick up the related bias and by that reveal a possible positive relationship between female education and growth.

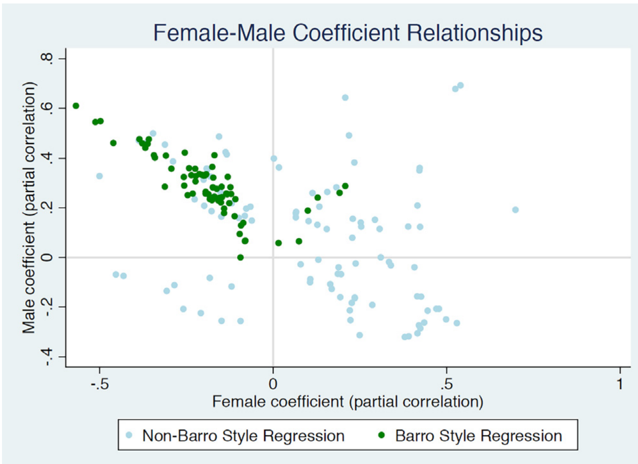


Fig. 2. Coefficient relationships, all estimates – Barro-specifications vs. non-Barro-specifications. Note: The green and light blue dots show the pair of male and female partial correlation coefficients between education and growth for Barro and non-Barro style regressions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

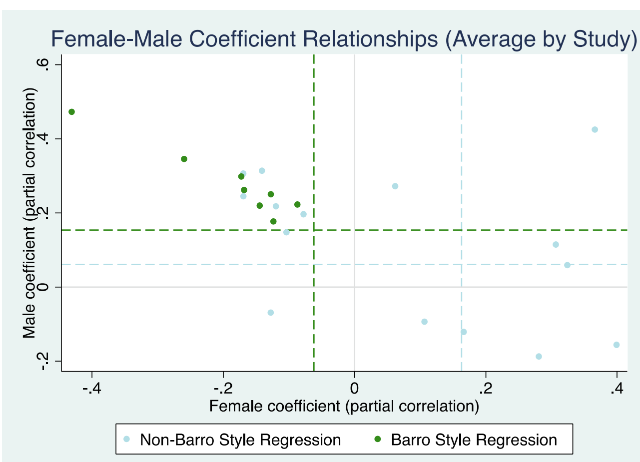


Fig. 3. Coefficient relationships, averaged by study – Barro-specifications vs. non-Barro-specifications. Note: The green and light blue dashed lines in Fig. 3 additionally represent the precision effect estimates controlling for the squared standard errors (PEESE) of the male and female effects for the Barro-style regressions and non-Barro style regressions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

negative (PEESE: −0.062, not significant) coefficients. Excluding the Barro-style estimates, we observe a relatively scattered picture across the remaining specifications. As in the previous section, the coefficient sizes and signs vary notably with different covariates and methodology. However, when looking at the PEESE-based correlations represented by the blue dashed lines we observe positive associations for both male and female education variables with economic growth. The PEESE weighted average is 0.163 (p-value <0.1) for the female coefficients and 0.061 (not significant) for the male coefficients. Thus, if we were to discount the findings using the Barro-style regressions for the reasons discussed above, the other studies suggest that female education is positively related to economic growth while male education is not, implying that reducing gender inequality is positively associated with economic growth. In the next section, we report the results of the MRA for the gap-studies.

Table 4
Gap-studies – Descriptive summary of results

Indicator	Share of coefficients	
	Positive, significant	Negative, significant
Female-to-male ratio (*F/M*)	0.8	0.025
Female-to-male logged difference (1 – (log *M* – log *F*))	0.5	0

Notes: Total number of *F/M*-estimates is 212 and total number of (log *M* – log *F*)-estimates is 4. *Significant* refers to statistical significance at least at the 10 percent level (p-value < 0.1).

7. Meta-regression analysis of female-male education gap and growth ("*Gap* studies)

A total of seventeen studies, including 216 relevant estimates, represent educational inequality by gender measured as a "gap" variable (e.g. a ratio of female education over male education). The use of a gender gap variable, relative to two separate indicators analyzed in section 6, has two advantages: First, it allows for a direct estimation of average partial correlation coefficient between estimates of the gender gap in education and economic growth. Second, it helps to avoid the problem of multicollinearity, which arises when female and male education variables are included in the same regression individually. To circumvent the latter, many studies choose to include a covariate for average education alongside with the gender gap measure, where the correlation between those two education variables is much lower compared to the studies which include education by gender separately (e.g. Klasen, 2002).

7.1. Descriptive evidence

Most estimates investigated here, 212 out of 216, include a female-male education *ratio* to measure the gender gap, while four regressions use log *difference* in male and female education. The gap-variable implies that an increase in this variable represents an increase in female relative to male education. The descriptive results reported in Table 4 suggest a positive relationship between gender equality in education and economic growth: Out of 212 regressions, in 80 percent of the cases the respective coefficient is positive and statistically significant at the conventional level; in only 2.5% of the cases it is negative and statistically significant.[17] Further, in two out of the four estimates which measure the gap variable as a logged difference (log *M* – log *F*) that we convert to the female-over-male coefficient for the final analysis, the correlation is positive and statistically significant at the conventional level. The first assessment of the pooled partial correlation coefficient (as described in Eq. 1) confirms that, by and large, lower gender inequality in education is positively associated with economic growth: The average partial correlation between the coefficient of the educational gender gap and growth is 0.21. Yet, heterogeneity in coefficients is large – ranging from negative 0.39 to positive 0.82 – with an average standard error of 0.10, ranging from 0.03 to 0.22.

7.2. Meta-analysis and assessment of publication bias

In Table 5 we report the average partial correlation coefficient between educational gender gap and economic growth using several standard meta-analysis techniques, as described in section 5. All models estimate standard errors clustered at the study-level. Column 1 in Table 5 displays the average partial correlation coeffi-

[17] Some regressions use the reverse ratio (i.e. a male-to-female ratio *M/F*) or reverse logged difference (i.e. log *M* – log *F*). For simplicity we have counted them towards the statistics in row two and four of Table 4 if *M/F* < 0 or (log *M* – log *F*) < 0, respectively, as well as p-value < 0.1.

Table 5
Average partial correlation of the educational gender-gap with growth.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Partial correlation coefficient									
	OLS	WLS	FAT-PET	PEESE	FAT-PET	PEESE	REML	REML	REML	REML
					Without outliers				Without outliers	
Constant	0.214***	0.221***	0.235	0.231**	0.151	0.185**	0.258***	0.258**	0.208***	0.208*
	(0.027)	(0.048)	(0.156)	(0.094)	(0.097)	(0.068)	(0.037)	(0.118)	(0.033)	(0.102)
SE			−0.190		0.680		−0.460	−0.460	0.059	0.059
			(1.531)		(0.886)		(0.387)	(1.119)	(0.348)	(0.930)
SE^2				−1.567		2.690				
				(7.685)		(5.127)				
Weights		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random effects							Yes	Yes	Yes	Yes
Small cluster adj.								Yes		Yes
No. of studies	17	17	17	17	17	17	17	17	17	17
Observations	216	216	216	216	212	212	216	216	212	212

Note: Standard errors reported in parentheses are clustered at the study level. *Constant* shows the average partial correlation of the gender gap in education with economic growth. Weights are equal to the inverse variance ($1/SE^2$). FAT-PET and PEESE test for publication bias.
* p-value < 10%.
** p-value < 5%.
*** p-value < 1%.

cient using a fixed effects meta-analysis model, i.e. a simple OLS estimation without weights, as described in Eq. (2).[18] In column 2, we adjust this model using weights of inverse variance (a weighted OLS), i.e. giving more weight to those estimates that are more precisely measured. Both specifications suggest a positive and significant correlation between the educational gender-gap and economic growth, ranging from 0.21 to 0.22, which represents a moderate size (Doucouliagos, 2011).[19] Yet, as described above, it is a reasonable assumption that publication bias and outliers may affect these estimates of the true underlying correlation.

Graphical assessment of the distribution of the estimates can give a first impression of whether these two concerns are relevant in our sample. The *funnel plot* in Fig. 4 shows the distribution of all estimates, plotting each partial correlation coefficient against a precision indicator, i.e. the inverse of the respective standard error (Iršová & Havránek, 2013; Stanley & Doucouliagos. 2012). The red line represents the weighted average partial correlation coefficient across studies, as specified in the model in Table 5, column 2. An unbiased funnel plot looks like a triangle that is symmetric around the true size of the correlation, while asymmetries may signal publication bias or outliers. The funnel plot in Fig. 4 shows that there are no strong asymmetries surrounding the average partial correlation coefficient (nor around 0). But the funnel plot does not show a very strong triangular shape and there appear to be some outliers among high precision estimates. Therefore, we assess publication bias more formally.

To assess publication bias formally, we apply the FAT-PET-PEESE strategy. We report the FAT-PET test for publication bias in column 3 of Table 5. We augment our meta-analysis by including the standard error (*SE*) of the partial correlation coefficient as an explanatory variable. To clarify, FAT-PET addresses the publication bias by controlling for the high correlation between small standard errors and availability (publication) of the study. The result in column 3 shows that the coefficient of *SE* is not statistically significant, hence we conclude that the coefficients in the sample do not suffer from publication bias. At the same time, the

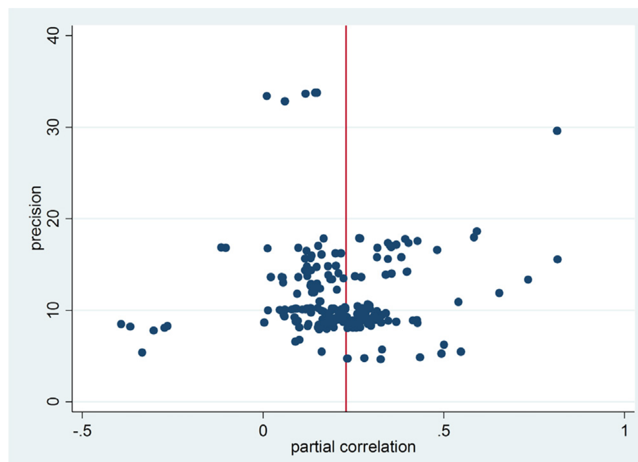


Fig. 4. Funnel plot.

size of the average partial correlation coefficient is robust to this adjustment. However, the coefficient loses statistical significance at the conventional levels (p-value = 0.15). As the FAT-PET method tends to underestimate a possible true underlying correlation, we conduct a second test (PEESE), which tends to perform better (if a non-zero effect exists). To carry out PEESE we replace *SE* with the squared standard errors (SE^2) in column 4. Again, we find a negative but statistically insignificant coefficient for SE^2, indicating that the test fails to reject the null hypothesis of no publication bias. This estimate represents the underlying genuine correlation between educational gender equality and growth, which is robust in size and statistically significant at the five percent level.[20]

Outliers. The funnel plot in Fig. 4 shows that there are a few estimates that might be outliers in our sample. As the FAT-PET-PEESE test can also be affected by outliers, we run the test for publication bias additionally without outliers. We follow Gallet and Doucouliagos (2017) and exclude outliers based on a rule of thumb: if the estimated standard deviation is larger than 3.5 then it is categorized as an outlier. As shown in column 5 and 6, the previously described test results are overall robust to this alteration,

[18] Please note since it is a partial correlation coefficient we cannot make a statement about the direction of causality. This is partially addressed by weighting the regression, giving more weights to those estimates that have smaller variance, i.e., standard errors.

[19] Doucouliagos (2011) suggests partial correlation coefficients of an absolute value between 0.07 and 0.17 to be considered as small, between 0.17 and 0.33 as moderate, and above 0.33 as large.

[20] Stanley (2008) notes that if FAT-PET fails to find a significant genuine average correlation PEESE should not be used. Stanley (2017) recommends to test the H_0: $\beta_0 \leq 0$ at the 10% level in the FAT-PET model to decide which model to employ.

Table 6
Description of regression characteristics.

Characteristic	Mean	SD	Definition
FLFP	0.10	0.30	Dummy = 1 if regression equation includes a control variable for female labor force participation; and 0 otherwise.
Fixed effects	0.12	0.32	Dummy = 1 if regression equation includes country or region dummies or country level panel fixed effects; and 0 otherwise.
Share of female authors	0.27	0.33	Continuous variable [0,1], indicates the study's share of female authors of total authors.
Published	0.49	0.50	Dummy = 1 if a study is published in a peer-reviewed international journal, and 0 if it is a published as working paper.
Economic controls	0.78	0.41	Dummy = 1 if regression equation includes control variables for openness, natural resources such as oil, landlocked, government expenditure, terms of trade, black market premium, inflation, money supply, agriculture value added, PPP, income inequality, GINI, financial sector, remittances, FDI, urbanization, or tax rate; and 0 otherwise.
Initial education	0.51	0.50	Dummy = 1 if a regression control for initial level of education in a country; and 0 otherwise.
Social/Institutional controls	0.56	0.50	Dummy = 1 if regression equation includes control variables for democracy, rule of law, language and religion, ethnic fractionalization, revolutions, assassinations, war, investment uncertainty, or gender wage gap; and 0 otherwise.
Enrollment	0.63	0.48	Dummy = 1 if education is measured in terms of enrollment (male, female or both); and 0 otherwise.
Dep. var.: Levels	0.40	0.49	Dummy = 1 if the dependent variable (GDP) is in levels; and 0 if the dependent variable is expressed as a change of GDP (growth).
Dep. var.: Logs	0.34	0.48	Dummy = 1 if the dependent variable is in logs; and 0 otherwise.
Source: Barro	0.35	0.48	Dummy = 1 if the education data is from Barro and Lee (1993, 1996, 2001, 2013); and 0 otherwise.

i.e. publication bias is not a strong concern in our sample. Yet, removing outliers does reduce the size of the average partial correlation coefficient, while the coefficient of the publication bias indicator changes signs but remains statistically insignificant.

Random effects. Note that the specifications discussed so far (columns 1–6) control only for the within-study variance and assume that between-study differences are random, i.e. the weighted OLS regressions are equivalent to *Fixed Effects MRA* (Stanley & Doucouliagos 2015, 2016). In other words, we assume that there is a single underlying size of the partial correlation coefficient, which is true for all the samples and years of all studies in the meta-analysis. However, this may not necessarily be the case as the studies in our sample are different in terms sample of countries, measures, methods and data sets used. Thus, the true correlation coefficient may vary between studies, i.e. the size could be higher or lower. Specifically, the random effects model assumes that the underlying correlation size in the seventeen studies included in our MRA are a random sample from a relevant distribution of correlation sizes, while the model estimates the mean of this distribution (Borenstein et al., 2010). Therefore, we run the MRA using a *Random Effects* approach to allow for the true correlation size to vary between studies and report the results in columns 7 to 10, which are comparable to those previously discussed. We do not find any statistical significance for publication bias at the conventional levels, outliers upwardly bias the mean estimate of the underlying true correlation size, while the correlation between the gender equality (*F/M*) in education and growth remains positive, sizable and statistically significant at the five percent level.

In sum, we find that the average correlation size is quite robust to different specifications and weights, that outliers matter, and that there is little evidence of publication bias. In our heterogeneity analysis below, we continue to report conservative estimates controlling for *SE* and compare our results with and without outliers, as well as with and without random effects.

7.3. Heterogeneity

As previously discussed, the coefficients included in our sample originate from regression equations that differ substantially in terms of datasets, methods, measure for education and income, and covariates used, etc. Table 6 quantifies the most important differences. These regression and study characteristics also likely affect the reliability of the estimates. For example, a larger set of control variables is likely to reduce left-out variable bias and the use of fixed effects can address time-invariant unobserved heterogeneity. In this section we investigate how these differing characteristics moderate the size of the average partial correlation estimate using fixed effects as well as random effects MRA models. Due to the limited number of degrees of freedom, we cannot include moderators for all possible study characteristics. Therefore, we restrict analysis to those that we regard as the most relevant ones, and combine some characteristics. Table 6 reports these characteristics and, among other things, it shows that 27 percent of authors in the included studies are female, suggesting that the share of female authors in our studies does not differ greatly from the share of female academics in economics.[21] About half of the estimates originate from studies that are published in international peer-reviewed journal, and the other half from studies published as working or discussion papers. Seventy-eight percent of estimates come from regression equations that use economic variables as covariates, 56 percent include control variables for social and institutional variables. A total of 63 percent of regressions measure education in terms of school enrollment, while 37 percent use other measures of education such as schooling attainment or literacy rates.

We explore heterogeneity in the average partial correlation coefficient by expanding our analysis from Table 5 with moderator variables from Table 6 in two sets. The first set, in Table 7, deals with model specification issues such as type of controls and methods used as well as external factors that might be related to finding positive and statistically significant estimates, such as publication status and perhaps the share of female authors. The second set, in Table 8, deals with measurement issues, for example how the education variable is measured, whether the estimation is in logs or in levels, the source of data, etc.

Table 7 column 1 shows that the size of the average partial correlation coefficient (the constant term) decreases when estimates originate from regression equations that include country fixed effects, i.e. control for unobserved time-invariant country specific characteristics. We also find that, a larger share of female authors is associated with a smaller average partial correlation coefficient. Inclusion of variables for female labor force participation (FLFP), initial education, and social and institutional controls increase the average partial correlation coefficient. Publication status does not significantly alter the size of the average partial correlation coefficient. These results are robust to removing outliers, except for the coefficient on FLFP, which no longer seems to have a moderating role (see column 4). Turning to the random effects model in

[21] In most OECD countries, women make up about 10–30% of professors in economics; the female share is higher at the doctoral or post-doctoral level. See e.g. Romero (2013).

Table 7
Effect size estimates, including specification-related moderators

	(1)	(2)	(3)	(4) Without outliers	(5)
	Partial correlation coefficient				
Constant	0.207** (0.071)	0.272*** (0.042)	0.272** (0.094)	0.158*** (0.045)	0.227*** (0.077)
SE	−0.662(1.013)	−0.651* (0.365)	−0.651(0.782)	0.072(0.671)	−0.184(0.635)
FLFP	0.091** (0.042)	0.043(0.035)	0.043(0.042)	0.029(0.034)	0.027(0.033)
Fixed effects	−0.099** (0.041)	−0.079*** (0.030)	−0.079* (0.041)	−0.068* (0.035)	−0.065* (0.035)
Share of female authors	−0.162* (0.081)	−0.164*** (0.039)	−0.164** (0.071)	−0.135* (0.068)	−0.140** (0.059)
Published	−0.016(0.041)	−0.025(0.031)	−0.025(0.040)	−0.019(0.037)	−0.028(0.035)
Economic controls	−0.072(0.076)	−0.134*** (0.034)	−0.134** (0.058)	−0.075(0.053)	−0.119** (0.046)
Initial education	0.228*** (0.060)	0.228*** (0.035)	0.228*** (0.070)	0.200*** (0.053)	0.202*** (0.058)
Social/Institutional controls	0.114*** (0.034)	0.098*** (0.026)	0.098** (0.034)	0.102*** (0.028)	0.091*** (0.029)
Weights	Yes	Yes	Yes	Yes	Yes
Random effects		Yes	Yes		Yes
Small cluster adj.			Yes		Yes
No. of studies	17	17	17	17	17
(Adj.) R^2	0.405	0.352	0.750	0.375	0.788
Observations	216	216	216	212	212

Notes: Standard errors in parentheses are clustered at the study level. In columns 4 and 5 we exclude outliers defined as those estimates which lie beyond an absolute standard deviation larger than 3.5.
* p-value < 10%.
** p-value < 5%.
*** p-value < 1%.

Table 8
Effect size estimates, including measurement-related moderators.

	(1)	(2)	(3)	(4) Without outliers	(5)
	Partial correlation coefficient				
Constant	0.184(0.115)	0.152*** (0.054)	0.152 (0.111)	0.097 (0.070)	0.094 (0.081)
SE	−0.564 (0.989)	−0.078 (0.418)	−0.086 (0.971)	0.404 (0.577)	0.486 (0.703)
Enrollment	0.117** (0.042)	0.097*** (0.030)	0.097*** (0.033)	0.088*** (0.030)	0.083*** (0.027)
Dep. var.: Levels	0.056 (0.067)	0.029 (0.027)	0.030 (0.076)	0.061 (0.052)	0.049 (0.058)
Dep. var.: Logs	−0.061 (0.042)	−0.020 (0.024)	−0.021 (0.044)	−0.042 (0.035)	−0.024 (0.036)
Source: Barro	0.009 (0.074)	0.004 (0.035)	0.005 (0.065)	0.028 (0.052)	0.021 (0.050)
Weights	Yes	Yes	Yes	Yes	Yes
Random effects		Yes	Yes		Yes
Small cluster adj.			Yes		Yes
No. of studies	17	17	17	17	17
(Adj.) R^2	0.220	0.128	0.675	0.204	0.728
Observations	216	216	216	212	212

Notes: Standard errors in parentheses are clustered at the study level. In column 4 and 5 we exclude outliers defined as those estimates which lie beyond an absolute standard deviation larger than 3.5.
** p-value < 5%.
*** p-value < 1%.

column 2, the signs of the moderator variables are similar to those of column 1, however there are some differences in terms of statistical significance. In particular, the inclusion of economic controls decreases the size of the average partial correlation coefficient, statistically significant at the one percent level. Overall, removing outliers yields similar results, while coefficients are generally smaller (column 5).

Next, in Table 8, the regressions related to measurement perform worse compared to Table 7 in terms of goodness of fit and significant effects. The only statistically significant finding is that the average partial correlation coefficient is larger when education is measured in terms of enrollment.

In Table 9 we include all covariates to see which of the moderators from Tables 7 and 8 remain significant when all are included. Column 1 reports the results for the fixed effects MRA and column 2 reports the results for the random effects MRA. Outliers are excluded. The specification-related moderators of Table 7 largely remain significant, while none from Table 8 do so. Specifically, in column 1 covariates that are significantly associated with the increase in the partial correlation coefficients are *Social/Institutional controls* and *Initial education* as before, while the estimate on *Enrollment* is not statistically significant anymore. The variables that are associated with decreasing the correlation size are *Fixed effects*, *Share of female authors* and *Economic controls* in both fixed effects and random effects regressions.

Based on this we can conclude that, on average, there is a statistically significant correlation between closing gender gap in education and economic growth. The size of the correlation is larger when models control for initial education levels in the country and include social/institutional controls in the regression analysis. Both types of control variables appear to be useful in reducing unobserved heterogeneity. The correlation size is smaller when the estimation uses country fixed effects and includes additional economic covariates, both of which are also suitable features to reduce left-out variable bias. If one considers specifications including these four desirable methodological features, one will have a larger than average partial correlation coefficient.

It is also notable that studies with a high female share in authorship are associated with smaller-sized correlations of the educational gender gap and growth – a result which is sizable and robust across all specifications. We tested whether this is due to a particular female author and found this not the case. There is a large body of literature on gender gaps in economics, including gender differences in the publication process (e.g. Hengel, 2017

Table 9
Effect size estimates, including all moderators.

	(1)	(2)
	Partial correlation (without outliers)	
Constant	0.163** (0.057)	0.153** (0.064)
SE	−0.101 (0.515)	0.162 (0.553)
FLFP	0.008 (0.023)	0.002 (0.027)
Fixed effects	−0.059*** (0.016)	−0.055*** (0.019)
Share of female authors	−0.131* (0.063)	−0.113* (0.059)
Published	0.010 (0.055)	0.023 (0.061)
Economic controls	−0.097** (0.039)	−0.101** (0.036)
Initial education	0.232*** (0.062)	0.212*** (0.063)
Social/Institutional controls	0.074** (0.027)	0.057** (0.027)
Enrollment	0.049 (0.029)	0.042 (0.031)
Dep. var.: Levels	−0.025 (0.054)	−0.017 (0.062)
Dep. var.: Logs	0.062 (0.039)	0.057 (0.043)
Source: Barro	−0.092 (0.065)	−0.092 (0.073)
Weights	Yes	Yes
Random effects		Yes
Small cluster adj.		Yes
No. of studies	17	17
(Adj.) R^2	0.474	0.817
Observations	212	212

Notes: Standard errors in parentheses are clustered at the study level. All columns exclude outliers defined as those estimates which lie beyond an absolute standard deviation larger than 3.5.
* p-value < 10%.
** p-value < 5%.
*** p-value < 1%.

and studies cited therein), and on the role of the sex of the experimenter in medical and other experiments (Chapman et al., 2018). But we have not found any other study that reported a relationship between female authors and empirical results using secondary data as we find here. If this is replicated in other studies, it clearly deserves further analysis.

In Appendix 3, we provide tests for robustness where we adjust standard errors by cluster size and use an alternative effect size calculation, concluding that all our MRA findings remain qualitatively the same.

8. Risk of bias assessment of IV studies

The discussions and analyses above are based on all studies that meet our inclusion criteria regardless of their risk of bias, i.e., the risk that studies may overestimate or underestimate the true correlation size. Traditionally, only randomized trials were considered reliable enough to be included in systematic reviews and meta-analysis for the evaluation of the evidence on the true "effect size" (Higgins & Green, 2011). Arguably, for many important research questions, like ours in this review, this approach is, as discussed above, not appropriate because we would be left without any eligible study. Several recent articles suggest to extend meta-analysis to non-experimental studies, in particular quasi-experiments that can generate evidence of similar strength to randomized trials, if certain assumptions are met and can be applied to research questions for which randomized trials are not possible (Sterne et al., 2016; Bärnighausen et al., 2017a, 2017b). We follow the definition of quasi-experiments suggested by King et al. (1995), which states that a quasi-experiment is "an observational study with an exogenous explanatory variable [or treatment, or exposure] that the investigator does not control". Bärnighausen et al. (2017a, 2017b) describe the study types that fall under this definition and discuss their assumptions. The only such quasi-experimental study type that we observe in our sample is instrumental variable regression. In total we identify six instrumental variable studies in our sample that instrument our key education variables. The methodological description of one of the studies (Barro, 1996a) allows for more than one interpretation on whether the schooling variables were instrumented, which we nevertheless include and discuss here (see Table 10).

First, we conduct a risk of bias assessment for these six selected studies. Official risk of bias assessments for instrumental variable regressions are still under development (Sterne et al., 2016), therefore we design our own simple assessment, following the conceptual framework of Bärnighausen et al. (2017b). In particular, the following four dimensions were identified as relevant and assessed independently by each author of this article. The inconsistencies were resolved by discussion of the entire team. The four dimensions and the relevant scores are:

1. Discussion of endogeneity: none (0 points), some discussion (1 point), appropriate discussion (2 points).
2. First stage: not mentioned (0 points), reported but weak or no F-test shown (1 point), reported and strong enough (2 points).
3. Exclusion restriction: not mentioned (0 points), verbally justified (1 point), supported by falsification tests or particularly strong verbal justification (2 points).
4. Local Average Treatment Effect (LATE): not mentioned (0 points), mentioned that local effect may differ from average effect in population/sample (1 point), mentioned and evidence shown or discussed that local effect is likely to differ from average effect in population/sample (2 points).

According to our assessment, an average score of zero up to 0.5 implies high-risk of bias, an average score greater 0.5 and up to one implies medium risk of bias and a score greater than one and up to two implies low risk of bias. The results of the assessment are shown in Table 10. Out of the six studies, none is classified as low-risk of bias. Two studies are classified as high risk of bias and four as medium risk of bias. Klasen (2002) and Licumba et al. (2015) are the two gap studies with medium risk of bias, and Dollar and Gatti (1999) and Knowles et al. (2002) are the two comparative studies with medium risk of bias. Three of these studies were published more than 15 years ago when reporting standards on sources of endogeneity, first stages, and exclusion restrictions were not as exacting and comprehensive as they are now. If they were written today, they would likely be expected to report more comprehensively on these issues.

In addition to the risk of bias score, we also pay particular attention to the validity of instruments, which is the most crucial condition for a causal interpretation of results. Previous evidence shows that many instruments used in the economic growth literature are weak, invalid or both (Bazzi & Clemens, 2013), which makes it even more important to assess this carefully. In the case of these six

Table 10
Risk of bias assessment for quasi-experimental studies.

Study	Endogeneity discussion	First stage	Exclusion restriction	LATE	Overall
Barro (1996a)	1.5	0	0.5	0	0.50 (high risk)
Dollar and Gatti (1999)	1	1	1	0	0.75 (medium risk)
Klasen (2002)	1.5	1	1	0	0.875 (medium risk)
Knowles et al. (2002)	2	0	1.5	0	0.875 (medium risk)
Licumba et al. (2015)	1	1	1	0	0.75 (medium risk)
Szulga (2006)	1	0	0	0	0.25 (high risk)

Table 11
Partial correlation of studies with lowest risk of bias and full sample.

	Klasen (2002)	Licumba et al. (2015)	All gap studies (REML)	
Partial correlation	0.183	0.403	0.258	
Standard error	0.108	0.077	0.118	
	Dollar and Gatti (1999)	Knowles et al. (2002)	All comparative studies (PEESE)	Non-Barro studies (PEESE)
Partial correlation female	0.106	0.422	−0.062	0.163
Standard error	0.070	0.125	0.037	0.089
Partial correlation male	−0.100	−0.277	0.154	0.061
Standard error	0.070	0.132	0.043	0.087

Notes: In each cell of the two left columns, simple averages of the relevant IV specifications of the respective studies are reported. The right columns repeat previous results from Table 5 and section 6 for comparison.

studies, the main problem is not instrument relevance which is often discussed and tested, but the exclusion restriction.

Barro (1996a) is scored as having a high risk of bias. In terms of IV strategy, it uses 5-year lags as instruments for all the variables in the growth regressions including the ones on female and male schooling. By now it is well known that lags are very poor instruments of the same variables as the unobserved confounders in the error term are likely to be persistent over time (Murray 2006). Hence, validity is a clear problem here. In the case of Szulga (2006), multiple instruments are used for the gendered education variable but eventually he reports the investment ratio, life expectancy and terms of trade to perform the best. Yet, earlier studies have argued that these variables can have direct causal effect on economic growth (Cervellati & Sunde, 2011; Acemoglu & Johnson, 2007, Mendoza, 1997; Borensztein et al., 1998, Kormendi & Meguire, 1985), putting into question the exclusion restriction assumption.

Religion is used as instrument by Dollar and Gatti (1999) and Licumba et al. (2015).[22] Studies have shown that religion is associated with economic growth (Barro & McClearly, 2003, Scully 1988), however the channels are subject to some debate and credible evidence is mostly limited to Weber's hypothesis of the prostestant work ethic in European countries (Andersen et al., 2017, Becker & Woessmann, 2009, Cantoni, 2015). While the instrument is not directly questioned by other studies, Dollar and Gatti (1999) and Licumba et al. (2015) also do not do much to justify its validity.

Knowles et al. (2002) use various climate variables (highest temperature, changes in the temperature, rainfall, distance from the equator) as instruments for male and female education and make sure to control for other relevant factors in their regressions, which are embedded in careful theoretical modeling. Nevertheless, doubts remain that their regression framework is able to capture all other relevant channels through which rainfall and temperature may affect economic growth because Barrios et al. (2010) and Dell et al (2014) show that rainfall and temperature shocks can influence economic growth through other channels.

The instruments used by Klasen (2002) include initial fertility rate and its changes while also controlling for population growth – one of the channels. In terms of fertility changes a sizeable literature in the field of demographic transition and economic growth shows that fertility rates can influence economic growth through inequality and the general education level (De La Croix & Doepke, 2003) or directly (Galor & Weil, 1996; Brander & Dorwik, 1994), albeit the direction of the causal effect can be both ways (Herzer et al., 2012). There are fewer potential threats to the exclusion restriction for initial fertility. Strulik and Vollmer (2015) show that initial fertility does not predict the onset of the demographic transition and thus subsequent economic growth but that there is some diffusion of takeoff, which is also in line with predictions of unified growth theory (Galor, 2005). While Klasen (2002) does not provide detailed justification for the exclusion restriction (other than report on an overidentification restriction test which alone is insufficient as a test for the validity of the exclusion restriction), he controls for population and labor force growth, which reduces threats to the validity but does not eliminate those entirely.

To summarize, while all instruments used are open to question and the studies do not provide comprehensive evidence for instrument validity, the four studies with medium risk of bias are also the studies where the instruments used are least problematic.

In Table 11 we compare the partial correlation coefficients of these four studies with medium risk of bias with the results of the full sets of gap studies and comparative studies. The partial correlation coefficients of the two medium risk gap studies are within the same range as the other gap studies. The partial correlation coefficients of the two medium risk comparative studies differ quite a bit from the other comparative studies, which is mainly explained by the Barro-style studies in this category that we discussed in section 6; they are more similar to the non-Barro studies (see Table 11). In fact, the two best identified comparative studies (Knowles et al., 2002; Dollar and Gatti, 1999) draw a much clearer picture of the relationship between gender inequality in education and growth. Both studies find a negative and partly significant average partial correlation for male education and a positive and significant average partial correlation for female education.

Overall, the four quasi-experimental studies with medium risk of bias find that lower gender inequality in education is associated with higher economic growth. However, as discussed, the causal interpretation of this result has to be taken with some caution, because each study has some limitation with respect to the validity of the instruments used.

9. Concluding remarks

In this article we conduct a systematic review and meta-regression analysis to identify the overall evidence on the link between gender inequality in education and economic growth. While the raw search records hit above 1000 studies, we ended up with 34 eligible cross-country studies based on the inclusion criteria, which required cross-country regressions with per capita economic growth as the outcome variable and a gendered educational variable as the explanatory one. This rather small number of eligible studies suggests that the cross-country literature focusing on the relationship between gender inequality in education and economic growth is not large but with growth prospects.

The empirical question of interest is largely based on cross-country regressions and thus suffers from the identification problems inherent in such aggregate analyses. At the same time, better identified micro-based methods to study the impact of educational gender gaps can readily study effects on household incomes, but have difficulty linking their findings to aggregate economic performance (Klasen, 2018). Thus, we believe that the cross-country literature provides important evidence on this link, albeit most of the current studies do not allow for a causal interpretation of the relationship.

[22] Dollar and Gatti (1999) also include civil liberties together with religion but this instrument also suffers from similiar validity issues.

As a result of our meta-analyses, first, we find that the size and the sign of the average correlation between female and male education (as separate right-hand side variables) and economic growth heavily depends whether a Barro-style specification was used or not. We argue that a Barro-style specification is likely to suffer from time-invariant omitted variable bias, especially related to unmeasured regional differences in economic performance. When excluding these studies, we report positive and statistically significant correlation between female education estimates and economic growth, while the smaller positive correlation with male education estimates is not statistically significant.

Second, the meta-analysis results of the eligible studies that use the ratio of female over male education (gender gap) as the right-hand side variable, show on average a positive and statistically significant correlation between gender equality in education and growth. The size of the correlation is larger, if initial education levels and social/institutional measures are controlled for, and smaller if country/region fixed effects and economic controls are included or the share of female authors is larger.

Third, we assess the (risk of) bias in overestimating or underestimating the correlation size in the quasi-experimental studies and evaluate the merits of the instruments used. We find that these studies have either high or medium risk of bias. In terms of the validity of quasi-experimental design, we find some variation in the quality of identification strategies and instruments. While some approaches can be outright rejected, others have their merits. Nevertheless, there is not a single study with a bullet proof identification strategy and the justification provided for the validity of the instruments in all of them may not withstand the critical assessment based on current standards.

In sum, this article suggests that there is a statistically significant positive association between gender equality in education and economic growth. Establishing causality is difficult due to the complexity of the relationship. The best quasi-experimental evidence that exists is not free from risk of bias or concerns about the validity of the identification strategy. Time will tell if this evidence can be further improved or if this is already the best that can be done. Either way, the relationship of gender equality in education and economic growth continues to be an important area of research.

Disclosure statement

We gratefully acknowledge funding from the Growth and Economic Opportunities for Women (GrOW) initiative, a multi-funder partnership between the UK's Department for International Development, the Hewlett Foundation and the International Development Research Centre.

Acknowledgements

We thank Hugh Waddington, Chris Doucouliagos, an anonymous referee, the responsible editor Michael Clemens, participants at workshops and conferences in Delhi, Göttingen, IAFFE conference in Seoul, Development Economics and Policy Conference in Zurich for helpful comments and suggestions. We gratefully acknowledge funding from the Growth and Economic Opportunities for Women (GrOW) initiative, a multi-funder partnership between the UK's Department for International Development, the Hewlett Foundation and the International Development Research Centre.

Appendix 1. Time series studies

Time series studies relate a time series of gender gaps in education to a time series of economic performance, sometimes controlling for additional covariates. The econometric methods used are quite different from the studies we just discussed which is why we discuss them separately. Among the 17 eligible time series studies, Dauda (2012) and Dauda (2013) are cases of self-plagiarism, reducing the number of original studies to 16. Ceesay (2013) reports separate time series results for 18 different countries (Algeria, Cameroon, Ethiopia, Gambia, Greece, India, Indonesia, Iran, Italy, Japan, Kenya, Malawi, Malaysia, Mali, Nigeria, Pakistan, Rwanda, Spain). Fatima (2013) reports separate time series results for two countries (Pakistan, Sri Lanka) and the remaining 14 studies report time series results for a single country. Nine of the 14 single country studies focus on Pakistan, and among the remaining single country studies there is one each for India, Japan Nigeria, Sudan and Turkey.

Ceesay (2013) uses GDP per capita as outcome and the ratios of female to male enrolment in secondary and tertiary education as explanatory variables. Assuming that male enrolment is higher than female enrolment, a higher ratio implies less educational gender inequality. For each of the 18 different countries OLS regressions with and without additional control variables are shown for the period from 1980 to 2010, and no attempt is made to address the endogeneity in the relationship. In the specifications with control variables the following results are statistically significant: Tertiary enrollment ratios in Albania, Iran, Spain (all positive) and Kenya, Mali (both negative). Secondary enrollment ratios in Greece, Kenya, Malaysia, Mali (all negative) and Gambia (positive). Risk of bias is very high in all these specifications.

Study periods in the Pakistan studies range from 1963 to 2012. Among the Pakistan studies, Alam et al. (2010), Amir and Mehmood (2012) and Fatima (2011) use some education indicators in the regressions without properly defining them in the text. They are therefore not further discussed in this review. Akram et al. (2011), Chaudhry (2007), Qureshi et al. (2007) use ratios of female to male education indicators as explanatory variables, mostly enrollment ratios and in the case of Chaudhry (2007) also literacy ratios. Again, assuming that male enrolment is higher than female enrolment, which is true for Pakistan, a higher ratio implies less educational gender inequality. Akram et al. (2011) find positive and statistically significant coefficients for professional, technical and higher education enrolment ratios and insignificant coefficients for primary and secondary enrolment ratios using VAR and cointegration techniques. Results of Augmented Dickey Fuller (ADF) tests and Johansen cointegration tests are shown to support the specification. Chaudhry (2007) finds positive and statistically significant coefficients both for literary ratios and primary enrolment ratios in an OLS specification without further attempts to address endogeneity. Qureshi et al. (2007) find positive and statistically significant coefficients for primary enrollment ratios and enrollment ratios in arts and science colleges and a negative and statistically significant coefficient for middle school enrollment ratios. They also mention to study high school, professional college, secondary vocational and university enrollment ratios but do not show results anywhere. The above-mentioned coefficients are based on a system of equations GMM specification and Jarque Bera, White, Durbin Watson and Ramsey specification tests are shown.

Khan (2015) and Stengos and Aurangzeb (2008) include both male and female education indicators in their analyses but no ratios or differences. Khan (2015) constructs a human capital index for females and males, which includes both health and education. The human capital index for females has a positive and statistically significant long-run coefficient in an error correction model, whereas the corresponding coefficient for males is statistically insignificant. The short-run coefficients are both statistically insignificant. The results are supported by a broad range of specification and cointegration tests. An additional Granger causality test shows that neither female nor male human capital Granger causes economic growth. Stengos and Aurangzeb (2008) find in

Granger causality tests that primary female enrollment, secondary male enrollment, development expenditure on secondary male education and development expenditure on secondary female education Granger cause economic growth. Results are not significant for male primary enrollment, female secondary enrollment and development expenditure on either female or male primary education. Results are not robust to the Levine and Renelt (1992) sensitivity analysis and, moreover, not much can be said about education inequality because effect sizes of male and female education are not investigated. Zaman et al. (2010) investigates only indicators of female education in Granger causality analysis. Technically, this study does not directly investigate the effect of educational gender inequality. But if one assumes that male education is closer to saturation whereas female education is at lower levels, then an increase in female education would imply a reduction in educational gender inequality. There is a unidirectional Granger causality from female primary and middle school enrollment to GDP and unidirectional Granger causality from GDP to female arts and science as well as university enrollment. There is no Granger causality for female high school, secondary vocational and professional college enrollment. Fatima (2013) includes both female and male years of schooling as well as the female to male ratio for years of schooling in a system of equations GMM model with the growth rate of GDP per capita as outcome. All three coefficients are statistically significant. The point estimate of male education is larger than that of female education (although no test for the difference is shown) and the ratio has a negative coefficient. Fatima (2013) also conducts the analysis for Sri Lanka and finds the same pattern as for Pakistan.

Awad et al. (2015) use an autoregressive distributed lag model (ARDL) and an error correction model to investigate the effect of female and male enrollment on GDP per capita in Sudan. The study period is 1960–2012. The ARDL coefficients are both statistically insignificant. In the error correction model, the short-run coefficients are both statistically insignificant. The long-run coefficients are both positive and statistically significant and have a very similar magnitude. The results are supported by a broad range of specification tests. Dauda (2012) investigates the effect of female and male secondary school enrollment on the growth rate of GDP in Nigeria between 1975 and 2008 in an error correction model. The coefficient of male secondary enrollment is statistically significant and positive, whereas the coefficient of female secondary enrollment is statistically insignificant. Results are supported by Johansen cointegration and ADF tests. Yumusak et al. (2013) conduct different cointegration tests of rate of girls among primary school graduates, high school graduates and university graduates with GDP growth for Turkey between 1968 and 2006. Puzzlingly, no cointegration analysis is conducted and just raw correlations of the three variables with GDP growth are shown. Self and Grabowski (2004) conduct Granger causality tests for the impact of female and male primary and secondary enrollment and change in female and male years of schooling on growth of GDP per capita in India between 1966 and 1996. They find that female and male primary and secondary enrollment all Granger cause growth of GDP. Not much can be said about the effect of educational inequality because effect sizes are not investigated. In terms of years of schooling only a change in female years of schooling but not in males Granger causes growth of GDP, therefore suggesting that reducing educational gender inequality has a positive effect on economic growth. Self and Grabowski (2005) investigate the effect of increases in years of female and male years of education (vocational and mainstream) on growth of GDP per capita in pre- and post-war Japan using vector error correction models. For the pre-war data they find that increases in both female and male years of mainstream education as well as female but not male vocational education are positively associated with growth of GDP per capita. For the post-war data they find that increases in both female and male years of vocational education as well as female but not male mainstream education had a positive correlation with growth of GDP per capita. Results are supported by ADF and Philips Perron tests. Unfortunately, no information is provided that would allow to compare female and male effect sizes.

Overall, the evidence from time series studies suggests that reducing gender inequality in education is associated with higher growth. We decided not to include the time series studies in the meta-analysis because of comparability issues, because of a high heterogeneity in methods, with many methods that have a high risk of bias. We also are concerned about external validity as all of these time series studies focus only on few countries, with more than half the studies being about Pakistan.

Appendix 2. Miscellaneous comparative studies

Here we summarize the three studies that run sub-national regressions using male and female education as covariates separately, and one Bayesian Model Averaging Study that also uses disaggregated education measures. One study investigates the relation of education gaps to growth in 75 Nepalese districts in 2001 (Dahal, 2012). Using OLS regressions, the study finds that female education has a larger positive and significant coefficient than male education (which itself is never significant) and that, additionally, a large education gender gap is associated with a reduction in GDP. Another study uses panel fixed effects regressions using annual data for India's states and finds that female literacy correlates with significantly higher income in 10 out of 14 specifications while male literacy is never significantly correlated with income levels (Esteve-Volart, 2004). A last study for 67 Turkish provinces using 5 year-averages from 1975 to 2000 show that both female and male education are related to GDP positively and significantly, but that only male education is associated with growth in less developed provinces, and female education in more developed ones (Tansel et al., 2013). To the extent one can generalize from these three countries, the results suggest that female education is more often associated with growth than male education, and thus indicates that reducing gender gaps in education would boost growth.

As the Bayesian model averaging study is also using the Barro-specification in a sample of only 50 countries from 1960 to 1996, it is not surprising that it finds that one of the 'robust' growth determinants in this particular sample (and given the particular choice of 94 possible growth determinants) is female years of tertiary schooling which is negatively correlated with growth (Abington, 2014).[23]

Appendix 3. Robustness tests

Adjusting standard errors for cluster size. Columns 8 and 10 of Table 5 as well as columns 3 and 5 of Tables 7 and 8, respectively, address a further issue related to the standard errors, which are clustered at the study level. A possible problem for our results is the uneven and small number of clusters, which may lead to an overestimation of the statistical significance of our coefficients. We follow Gallet and Doucouliagos (2017) to adjust the standard errors in our random effects model for cluster size. Our results

[23] One should also note that the 'robustness' of growth determinants using this method depends greatly on the sample and the covariates considered. For example, Abington (2014), show that their study has little overlap of robust growth determinants with an earlier study by Sala-i-Martin et al. (2004) even though all they do is to add some more human capital variables to the set of growth determinants.

remain qualitatively similar to the models with the non-adjusted standard errors.

Alternative effect size calculation. So far, we have estimated the true underlying correlation based on a partial correlation coefficients, assuming that the standard errors in our meta-regression equation are normally distributed. In case the latter assumption is questionable, Stanley and Doucouliagos (2012) suggest to test for the robustness of these results by implementing a Z-transformation of the partial correlation coefficients and the standard error. We visually inspect the standard errors and find them to be normally distributed; for robustness we, nevertheless, report results using the Z-transformed partial correlation coefficients and standard errors. The results of our main analysis based on this transformation are presented in Table A1. While the coefficients in all presented tests are decreasing compared to the results in Table 5, partly reducing the correlation size from moderate to small, our qualitative conclusions from above still hold.

Table A1
Average effect of education gender-gap on growth (Z-transformed partial correlation coefficient)

	(1)	(2)	(3)	(4)	(5)	(6)
	Z-transformed					
	FAT-PET	PEESE	REML	FAT-PET	PEESE	REML
	Without outliers					
Constant	0.139	0.180**	0.183***	0.126	0.167**	0.167***
	(0.103)	(0.075)	(0.042)	(0.086)	(0.064)	(0.035)
SE	0.942		0.435	0.937		0.478
	(0.960)		(0.437)	(0.794)		(0.370)
SE^2		4.404			4.537	
		(5.878)			(4.997)	
Weights	Yes	Yes	Yes	Yes	Yes	Yes
Random effects			Yes			Yes
No. of studies	17	17	17	17	17	17
Observations	216	216	216	213	213	213

Notes: Standard errors in parentheses are clustered at the study level. Columns 4 to 6 exclude outliers defined as those estimates which lie beyond an absolute standard deviation larger than 3.5.

** p-value < 5%.
*** p-value < 1%.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.worlddev.2019.05.006.

References

Acemoglu, D., & Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of Political Economy, 115*(6), 925–985.
Andersen, T. B., Bentzen, J., Dalgaard, C. J., & Sharp, P. (2017). Pre-reformation roots of the protestant ethic. *The Economic Journal, 127*(604), 1756–1793.
Bandiera, O., & Natraj, A. (2012). Does gender inequality hinder development and economic growth? Evidence and policy implications. *World Bank Research Observer, 28*, 2–21.
Bärnighausen, T., Tugwell, P., Røttingen, J. A., Shemilt, I., Rockers, P., Geldsetzer, P., et al. (2017). Quasi-experimental study designs series: Uses and value. *Journal of Clinical Epidemiology, 89*, 21–29.
Bärnighausen, T., Oldenburg, C., Tugwell, P., Bommer, C., Ebert, C., Barreto, M., et al. (2017). Quasi-experimental study designs series: Assessing the assumptions. *Journal of Clinical Epidemiology, 89*, 53–66.
Barro & Sala-i-Martin (1995). *Economic growth*. Cambridge: Harvard University Press.
Barro, R. (1991). Economic growth in a cross-section of countries. *Quarterly Journal of Economics, 106*(2), 407–433.
Barro, R., & Lee, J.-W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics, 104*, 184–198.
Barro, R., & Lee, J.-W. (1994). Sources of economic growth. *Carnegie-Rochester Conference Series on Public Policy, 40*, 1–46.
Barro, R. J., & Lee, J. W. (1996). International measures of schooling years and schooling quality. *The American Economic Review, 86*(2), 218–223.
Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics, 104*, 184–198.
Barro, R. J., & Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics, 32*(3), 363–394.
Barro, R. J., & Lee, J.-W. (2001). International data on educational attainment: updates and implications. *Oxford Economic Papers, 53*(3), 541–563.
Barro, R. J., & McCleary, R. M. (2003). Religion and economic growth across countries. *American Sociological Review, 68*(5), 760.
Barrios, S., Bertinelli, L., & Strobl, E. (2010). Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. *The Review of Economics and Statistics, 92*(2), 350–366.
Bazzi, S., & Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics, 5*(2), 152–186.
Becker, G. (1981). *A treatise on the family*. Chicago: University of Chicago Press.
Becker, S., & Woessmann, L. (2009). Was Weber wrong? A human capital theory of protestant economic history. *Quarterly Journal of Economics, 124*(2), 531–596.
Benos, N., & Zotou, S. (2014). Education and economic growth: A meta-regression analysis. *World Development, 64*, 669–689.
Bloom, D., & Williamson, J. (1998). Demographic transition and economic miracles in emerging Asia. *The World Bank Economic Review, 12*(3), 419–455.
Borenstein, M. et al. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111.
Borensztein, E., De Gregorio, J., & Lee, J. W. (1998). How does foreign direct investment affect economic growth? *Journal of International Economics, 45*(1), 115–135.
Brander, J. A., & Dowrick, S. (1994). The role of fertility and population in economic growth. *Journal of Population Economics, 7*(1), 1–25.
Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine, 20*(6), 825–840.
Busse, M., & Spielmann, C. (2006). Gender inequality and trade. *Review of International Economics, 14*(3), 362–1279.
Cantoni, D. (2015). The economic effects of the protestant reformation: Testing the Weber hypothesis in the German Lands. *Journal of the European Economic Association., 13*(4), 561–598.
Cavalcanti, T. V., & Tavares, J. (2016). The output costs of gender discrimination: A model-based macroeconomic estimate. *Economic Journal, 126*, 109–134.
Chapman, C., Benedict, C., & Schiöth, H. (2018). Experimenter gender and replicability in science. *Science Advances, 4*(1) e1701427.

Cervellati, M., & Sunde, U. (2011). Life expectancy and economic growth: the role of the demographic transition. *Journal of economic growth, 16*(2), 99–133.
Cuberes D. and Teignier M. (2016). Aggregate effects of gender gaps in the labor market: A quantitative estimate. *Journal of Human Capital, 10*(1), 1–32.
De La Croix, D., & Doepke, M. (2003). Inequality and growth: why differential fertility matters. *American Economic Review, 93*(4), 1091–1113.
Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature, 52*(3), 740–798.
Doucouliagos, C. (2011). How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Economics Series Working Paper 05, Deakin University.
Duflo, E. (2012). Women's empowerment and economic development. *Journal of Economic Literature, 50*(4), 1051–1079.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.
Egger, M. et al. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634.
Gallet, C. A., & Doucouliagos, H. (2017). The impact of healthcare spending on health outcomes: A meta-regression analysis. *Social Science & Medicine, 179*, 9–17.
Galor, O., (2005). From stagnation to growth: unified growth theory. In Aghion, P., &, Durlauf, S.N., (eds.) Handbook of economic growth. North-Holland: Amsterdam.
Galor and Weil (1996). The gender gap, fertility, and growth. *American Economic Review, 86*, 374–387.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
Hengel, E. (2017). *Publishing while Female. Are women held to higher standards? Evidence from peer review*. Faculty of Economics: University of Cambridge.
Higgins, J., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions. Version 5.1*. Oxford: The Cochrane Collaboration.
Herzer, D., Strulik, H., & Vollmer, S. (2012). The long-run determinants of fertility: one century of demographic change 1900–1999. *Journal of Economic Growth, 17* (4), 357–385.
Iršová, Z., & Havránek, T. (2013). Determinants of horizontal spillovers from FDI: Evidence from a large meta-analysis. *World Development, 42*, 1–15.
King, G., Keohane, R. O., & Verba, S. (1995). The importance of research design in political science. *American Political Science Review, 89*(2), 475–481.
King, E. S. Klasen, & Porter, M., (2009). Women and Development. In Lomborg, B. (ed.) Global crises, global solutions, (2nd ed.) Cambridge: Cambridge University Press.
Klasen, S. (2006). Gender and pro-poor growth. In L. Menkoff (Ed.), *Pro-poor growth: Evidence and policies*. Berlin: Dunker&Humblot.
Klasen, S. (2016). Gender, institutions, and economic development. Courant Research Center Discussion Paper.
Kormendi, R. C., & Meguire, P. G. (1985). Macroeconomic determinants of growth: cross-country evidence. *Journal of Monetary economics, 16*(2), 141–163.
Lagerlöf, N. (2003). Gender equality and long-run growth. *Journal of Economic Growth, 8*(4), 403–426.
Levine, R., & Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *The American Economic Review*, 942–963.
Lorgelly, P., & Owen, D. (1999). The effect of female and male schooling on economic growth in the Barro-Lee model. *Empirical Economics, 24*(3), 537–557.
Mendoza, E. G. (1997). Terms-of-trade uncertainty and economic growth. *Journal of Development Economics, 54*(2), 323–356.
McKinsey (2015). *The power of parity: How advancing women's equality can add $12 trillion to global growth*. McKinsey Global Institute.
Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives, 20*(4), 111–132.
Murthi, M., Guio, A.-C., & Drèze, J. (1995). Mortality, fertility, and gender bias in India: A district-level analysis. *Population and Development Review, 21*(4), 745–782.
OECD (1995). *Household production in OECD countries*. Paris: OECD.
Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing.
Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review, 15*(3), 367–391.
Romero, J. (2013). Where are the Women? *Econ Focus, 7*(2), 12.
Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*, 135–158.
Sala-i-Martin, X., Doppelhofer, G., & Miller, R.I. (2004). Determinants of long-term growth: A Bayesian Averaging of Classical Estimates (BACE) approach. American Economic Review, 94(4), 813–835.
Scully, G. W. (1988). The institutional framework and economic development. *Journal of Political Economy, 96*(3), 652–662.
Schultz, T. P. (2002). Why governments should invest more to educate girls. *World Development, 30*(2), 207–225.
Seguino, S. (2000a). Gender inequality and economic growth: A cross-country analysis. *World Development, 28*, 1211–1230.
Seguino, S. (2000b). Accounting for gender in Asian economic growth. *Feminist Economics, Taylow and Francis Journals, 6*(3), 27–58.
Stanley, T. D., & Jarrell, S. B. (1989). Meta-Regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys, 3*(2), 161–170.
Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *The Journal of Economic Perspectives, 15*(3), 131–150.
Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys, 19*(3), 309–345.
Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection. *Oxford Bulletin of Economics and Statistics, 70*(1), 103–127.
Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science, 8*(5), 581–591.
Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. Oxford: Routledge.
Stanley, T. D., & Doucouliagos, H. (2015). Neither fixed nor random: weighted least squares meta-analysis. *Statistics in Medicine, 34*(13), 2116–2127.
Stanley,T.D., & Doucouliagos, H. (2016). Neither fixed nor random: weighted least squares meta regression analysis. Res. Synthesis Methods. In press.
Sterne, J. A. et al. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ, 355* i4919.
Strulik, H., & Vollmer, S. (2015). The fertility transition around the World. *Journal of Population Economics, 28*(1), 31–44.
Summers, L. (1994). *Investing in all the people*. Washington DC: The World Bank.
Taylor, A. (1998). On the costs of inward looking development. *Journal of Economic History, 58*(1), 1–28.
Tertilt, M., & M. Doepke. (2014). Does female empowerment promote economic development? NBER Working Paper No. 19888.
Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine, 18*(20), 2693–2708.
UN. (2008). The 2008 System of National Accounts. New York: UN.
UNDP (1995). *Human development report*. New York: Oxford University Press.
Waddington et al. (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of Development Effectiveness*, 359–387.
World Bank (1994). *The East Asian miracle*. Washington DC: The World Bank.
World Bank (2001). *Engendering development*. Washington DC: The World Bank.
Zürcher, C. (2017). What do we (not) know about development aid and violence? A systematic review. *World Development, 98*, 506–522.

References of eligible studies

Abington, C. R. (2014). Determining the importance of human capital for growth using a Bayesian averaging of classical estimates approach. *Global Business and Economics Research Journal, 3*(4), 62–78.
Akram, N., Hamid, A., & Bashir, S. (2011). Gender differentials in education and their impact on economic growth of Pakistan. *Journal of Business & Economics, 3*(1), 102.
Akyeampong, E., & Fofack, H. (2013). The contribution of African Women to economic growth and development in post-Colonial Africa. World Bank Policy Research Working Paper, 6537.
Alam, T., Ellahi, N., Bukhari, M. A., & Jamil, N. (2010). Female education and economic performance: a time series analysis for Pakistan. Journal of Education and Sociology.
Amin, M., Kuntchev, V., & Schmidt, M. (2015). Gender inequality and growth: the case of rich vs. poor countries. Policy Research Working Paper Series.
Amir, M., & Mehmood, B. (2012). How gender inequalities affect economic growth in Pakistan: A cointegration analysis. *Actual Problems of Economics, 135*, 223–229.
Awad, A., Yussof, I., Sarmidi, T., & Ismail, R. (2015). Economic growth and human development-does gender matter? *African Journal of Economic and Sustainable Development, 4*(1), 74–108.
Baliamoune-Lutz, M., & McGillivray, M. (2009). Does gender inequality reduce growth in Sub-Saharan African and Arab countries? *African Development Review, 21*(2), 224–242.
BaliamouneLutz, M., & McGillivray, M. (2015). The impact of gender inequality in education on income in Africa and the Middle East. *Economic Modelling, 47*, 1–11.
Bandara, A. (2015). The economic cost of gender gaps in effective labor: Africa's missing growth reserve. *Feminist Economics, 21*(2), 162–186.
Barro, R., & Lee, J. (1994). Sources of economic growth. *Carnegie-Rochester Conference Series on Public Policy*.
Barro, R. (1996a). Inflation and growth. Review, Federal Reserve Bank of St. Louis 78(3).
Barro, R. (1996). Democracy and growth. *Journal of Economic Growth, 1*, 1–27.
Bazzi, S., & Clemens, M. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics, 5*(2), 152–186.
Benavot, A. (1989). Education, gender, and economic development: A cross-national study. *Sociology of Education*, 14–32.
Brummet, Q. (2008). The effect of gender inequality on growth: a cross-country empirical study. *The Park Place Economist, 16*(1), 13–23.
Caselli, F., Esquivel, G., & Lefort, F. (1996). Reopening the convergence debate: a new look at cross-country growth empirics. *Journal of Economic Growth, 1*(3), 363–389.
Ceesay, E. K. (2013). *The effects of gender inequality in education (human capita proxy), employment (labor market participation proxy) and its impacts on economic growth*. Inequality and Growth.
Chaudhry, I. (2007). *Gender inequality in education and economic growth: case study of Pakistan*. Pakistan Horizon.

Cooray, A., & Mallick, S. (2011). *What explains cross-country growth in South Asia? Female education and the growth effect of international openness*. Brooks World Poverty Institute Working Paper Series.
Cooray, A., Mallick, S., & Dutta, N. (2014). Gender-specific human capital, openness and growth: Exploring the linkages for South Asia. *Review of Development Economics, 18*(1), 107–122.
Dahal, M. P. (2012). Effect of gender gap in education on district level GDP per capital of Nepal. *Economic Journal of Development Issues, 13*, 65–74.
Dauda, R. O. S. (2012). Female education for growth and structural transformation in Nigeria: An empirical investigation. *Nigerian Journal of Economic and Social Studies, 54*(2), 233–252.
Dauda, R. O. S. (2013). Does female education promote economic performance? Evidence from Nigeria. *International Journal of Economics and Finance, 5*(1), 201–209.
Dollar, D., & Gatti, R. (1999). Gender inequality, income, and growth: are good times good for women? Policy Research Report Discussion Paper No. 1. Development Research Group, The World Bank Washington, DC.
El Alaoui, A. (n.d.). Impact of women's education on the economic growth: An empirical analysis applied to Morocco, Algeria, Tunisia, and Egypt. MPRA Paper.
Esteve-Volart, B. (2004). *Gender discrimination and growth: Theory and evidence from India*. Mimeographed: University of York, Canada.
Fatima, G. (2011). Female education as a determinant of economic growth: The case study of Pakistan. *Contemporary Issues in Education Research (Online), 4*(11), 15.
Fatima, G. (2013). Gender inequality in human capital accumulation and economic growth: A comparative analysis of Pakistan and Sri Lanka. *Asia Pacific Journal of Social Work and Development, 23*(4), 242–252.
Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review*, 869–887.
Hassan, G. M., & Cooray, A. (2015). Effects of male and female education on economic growth: Some evidence from Asia. *Journal of Asian Economics, 36*(C), 97–109.
Hassan, R., & Mirza, M. S. (2005). Education inequality and economic growth: FRAMEWORK for the evaluation of Pakistan's education policy. *MPRA Paper*.
Hill, M. A., & King, E. (1995). Women's education and economic well-being. *Feminist Economics, 1*(2), 21–46.
Huffman, W. E., & Orazem, P. F. (2004). The Role of agriculture and human capital in economic growth: Farmers, schooling and health. Handbook of Agricultural Economic, 101.
Kalaitzidakis, P., Mamuneas, T. P., Savvides, A., & Stengos, T. (2001). Measures of human capital and nonlinearities in economic growth. *Journal of Economic Growth, 6*(3), 229–254.
Khan, M. K. (2015). Contribution of female human capital in economic growth: an empirical analysis of Pakistan (19722012). *Quality & Quantity, 50*(2), 709–728.
Klasen, S. (2002). Low schooling for girls, slower growth for all? Cross-country evidence on the effect of gender inequality in education on economic development. *World Bank Economic Review, 17*(2), 315–316.
Klasen, S., & Lamanna, F. (2008). The impact of gender inequality in education and employment on economic growth in developing countries: Updates and extensions. *Feminist Economics, 15*(3), 91–132.
Knowles, S., Lorgelly, P. P. K., & Owen, P. D. (2002). Are educational gender gaps a brake on economic development? Some cross-country empirical evidence. *Oxford Economic Papers, 54*(1), 118–149.
Licumba, E. A., Dzator, J., & Zhang, J. X. (2015). Gender equality in education and economic growth in selected Southern African countries. *The Journal of Developing Areas, 49*(6), 349–360.
Lorgelly, P. K., & Owen, P. D. (1999). The effect of female and male schooling on economic growth in the Barro-Lee model. *Empirical Economics, 24*(3), 537–557.
No, W. P., Ferrant, G., & Kolev, A. (2016). Does gender discrimination in social institutions matter for long-term growth ? Cross-country evidence. OECD Development Centre Working Papers, (330).
Perotti, R. (1996). Growth, income distribution, and democracy: what the data say. *Journal of Economic Growth, 1*(2), 149–187.
Petticrew, M., & Roberts, H. (2006). How to appraise the studies: an introduction to assessing study quality. Systematic Reviews in the Social Sciences: A Practical Guide, 125–163.
Qureshi, S. A., Khan, M., Ul Husnain, M. I., & Iqbal, N. (2007). Gender, environment, and sustainable economic growth [with comments]. *The Pakistan Development Review*, 883–894.
Sadeghi, J. M. (n.d.). The relationship of gender difference in education to economic growth: A cross-country analysis. Working Papers.
Seguino, S. (2000b). Accounting for gender in Asian economic growth. *Feminist Economics, 6*(3), 27–58.
Seguino, S. (2000c). Gender inequality and economic growth: A cross-country analysis. *World Development, 28*(7), 1211–1230.
Seguino, S. (n.d.). All types of inequality are not created equal: Divergent impacts of inequality on economic growth. Economics Working Paper Archive.
Self, S., & Grabowski, R. (2004). Does education at all levels cause growth? India, a case study. *Economics of Education Review, 23*(1), 47–55.
Self, S., & Grabowski, R. (2005). Gender, vocational education, and economic development: The Japanese experience. *Review of Applied Economics, 1*(1).
Stengos, T., & Aurangzeb, A. (2008). An empirical investigation of the relationship between education and growth in Pakistan. *International Economic Journal, 22*(3), 345–359.
Szulga, R. (2006). *Female labor force participation, education and economic development*. Davis: University of California.
Tansel, A., Gungor, N. D., Güngör, N. D., Gungor, N. D., & Güngör, N. D. (2013). Gender effects of education on economic development in Turkey. *Journal of Economic Studies, 40*(6), 794–821.
Thévenon, O., Ali, N., Adema, W., Pero, A. S. del, Thevenon, O., Ali, N., ... del Pero, A. (2012). Effects of reducing gender gaps in education and labour force participation on economic growth in the OECD. OECD Social, Employment and Migration Working Papers.
Yumusak, I. G., Bilen, M., & Ates, H. (2013). The impacts of gender inequality in education on economic growth in Turkey. *Procedia – Social and Behavioral Sciences, 103*, 1093–1103.
Zaman, K., Khan, M. M., Ahmad, M., & Ikram, W. (2010). Do female enrolment rates cause economic growth in Pakistan? *Asian Social Science, 6*(11), 256.