Courant Research Centre

'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 255

Educational Gender Gaps and Economic Growth: A Systematic Review and Meta-Regression Analysis

Anna Minasyan, Juliane Zenker, Stephan Klasen and Sebastian Vollmer

August 2018

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Educational Gender Gaps and Economic Growth: A Systematic Review and Meta-Regression Analysis[1]

Anna Minasyan*, University of Groningen
Juliane Zenker*, University of Göttingen
Stephan Klasen§, University of Göttingen
Sebastian Vollmer, University of Göttingen

August 2018

Keywords: Gender gaps, Education, Growth, Systematic review, Meta-analysis

JEL Codes: O47, I24, I25

Abstract

We conduct a systematic review and meta-analysis of the empirical literature on the impact of gender inequality in education on per capita economic growth, including cross-country, time series, and sub-national growth regressions. Studies using male and female education as separate covariates show a larger effect of female than male education on growth, except when an arguably problematic regression specification popularized by Barro and co-authors is used. We conduct a meta-regression analysis for studies that use the female-male ratio of education as explanatory variable. There we find evidence for a positive and statistically significant relationship between gender equality in education and growth based on 216 estimates from 17 such studies. We find that the average partial correlation coefficient between economic growth and the ratio of female over male education is 0.25, which is a moderate effect. The effect does not appear to be influenced by publication bias, it increases when one controls for initial education levels and social/institutional controls, while it falls with the use of fixed effects, the inclusion of economic controls, and the share of female authors.

[1]**Acknowledgements:**
We thank Hugh Waddington, Chris Doucouliagos, 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.

*First authors. §Correspondence to sklasen@uni-goettingen.de.

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). While some gender gaps are present in all countries of the world, gender gaps 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, also the development impact of gender inequality 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, and Dreze, 1995). Soon thereafter first studies appeared that investigated the impact of gender gaps on economic performance (e.g. Hill and 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 impact of gender gaps in employment, pay, health, laws, and empowerment on economic growth within this growth-regression framework, by far the largest number of studies has focused on the impact of gender differences in education on economic growth. This is partly related to the fact that human capital is 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 available and quite reliable metrics of education quantity by gender, including enrolment rates, literacy rates, and years of schooling by sex (e.g. Barro and Lee, 2013). Lastly, there have been some noted controversies on the impact of female education on economic growth. On the one hand, there are several theoretical mechanisms that suggest that gender gaps in education could promote economic growth, while there are quite a few mechanisms that suggest the opposite (see discussion below).

On the empirical side, Barro and Lee (1994) and Barro and Sala-i-Martin (1995) reported the 'puzzling' finding that more female years of schooling reduce economic growth, while the reverse was the case for males. 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 and Gatti, 1999; Klasen, 2002; Lorgelly and Owen, 1999, and Knowles, Lorgelly, and 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, therefore, aims to systematically assess the evidence and synthesize the differing and partly opposing findings (Stanley and Jarrell, 1989; Stanley, 2001). The large body of relevant literature consists of cross-country studies (including pure cross-sections or panels), single country time series studies, and single country cross-regional studies. We group the studies accordingly for our analysis. The cross-country studies can be further divided into *comparative* and *gap* 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 average effect sizes.[2] The *comparative* studies, using male and female education as separate covariates in the growth regression, show a larger effect of female than male education on growth, except when a regression specification popularized by Barro and co-authors (e.g. Barro and 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 based on 216 estimates from 17 such studies. 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 effect of educational gender equality on growth. Finally, the time series analyses we investigate are based on a few countries and generally weak methods. We, therefore, refrain from drawing strong generalized conclusions from this set of studies. Yet, also the evidence from time series studies suggests that reducing gender inequality in education may have a growth promoting effect.

2. Conceptual framework

There exist theoretical arguments that highlight both mechanisms for a positive as well as a negative effect of educational gender gaps on economic growth, however we are not aware of theoretical literature that compares the magnitudes of the different effect sizes to make statements about the net effect of the various mechanisms. Therefore, 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. The first goes back to Becker (1981), essentially arguing 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). Whatever the merits of the argument, it is likely to become less relevant as fertility falls and household production becomes less time-consuming (also due to improved technologies). A second argument can be made following suggestions by Tertilt and Doepke (2014): due to different

[2] Throughout the paper we make use of the terms *effect* and *effects sizes* to comply with the terminology common to the meta-analysis literature. However, most of the estimates in our research synthesis are based on regression equations that do not allow for a causal interpretation.

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 are a substantial number of papers arguing the reverse, i.e. that gender gaps in education reduce economic performance. As a first argument, the 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 and Gatti, 1999; Teignier and Cuberes, 2015). 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 this 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 has a positive impact on economic growth (World Bank 2001; King, Klasen, and 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 causes poverty among the next generation leading to low-income poverty traps (e.g. Galor and Weil 1996; Lagerlöf 2003). But there is also an important timing issue involved here. Reducing gender gaps in education will 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]

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, a strategy that is now finding followers in South Asia and individual countries across the developing world (e.g. Seguino, 2000a, b).[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 and Spielmann 2006).

Given the competing arguments, it becomes an empirical question whether and to what extent gender inequality has an impact on economic growth. As the different models suggest different mechanisms, ideally one would look into these mechanisms in the empirical

[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.
[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.

literature. Our meta-regression can partly address this question by examining the role of particular control variables – some of which represent mechanisms.

3. Systematic-review methodology

3.1 Criteria for the inclusion of studies

Following Petticrew and Roberts (2006), we use the PICOS model (population, intervention, comparison, outcome and setting) to define the inclusion criteria for our review.

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.

Intervention. We are looking at the effects of changes and levels of educational gender gap within a country for the largest time possible based on observational and macroeconomic data. On the right-hand side of the estimation equation must be either the levels of female and male education separately (both in one regression) or a measure of the gender gap in education. All educational indicators are considered (enrolment, attainment, years of schooling). Studies can also include instrumental variables for the educational gap as well as time lags of the gendered educational gap. We drop studies that include only male or only female education as they cannot be used to assess the impact of educational gender gaps on growth.

Comparison. We consider only quantitative, observational studies that include regression analyses that aim to evaluate the effect of educational gender gaps on the *outcome* specified below. We include studies that have a clearly defined sample, method and results description. Comparison is based on educational gender differences between countries as well as changes of the gap size within a country over time. Based on the research design, we categorize the studies into the following groups:

a) Within-country time series: These studies use time series econometric techniques to relate a time series of educational gender gaps to a time series of growth in a particular country. While these studies will be summarized in the systematic review, we do not include them in the meta-analysis.

b) Cross-country cross-sectional regression analysis: These studies use variation between countries.

c) Panel cross-country studies: These studies use variation across countries and 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. The outcome is economic growth defined by the growth rate of GDP per capita. In some cases, the outcome can also be the level of per capita income measures if the study design allows to translate this to economic growth. 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.

3.2 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 impact of gender inequality in education on the economic growth. 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.[6]

We have used four easily accessible research databases: EconLit, IDEAS, Web of Science and Google Scholar. The first two contain papers 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), and economic growth. We combine three to four sets of synonymous terms in several ways to capture all potentially relevant studies. See the Appendix for a detailed overview on all applied search strings and search specific results.[7] 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 appendix.

As detailed in Figure 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

[6] 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.

[7] 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*)[7] 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 appendix for details.

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.[8]

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 total, 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 3a) to d), above by two reviewers independently.

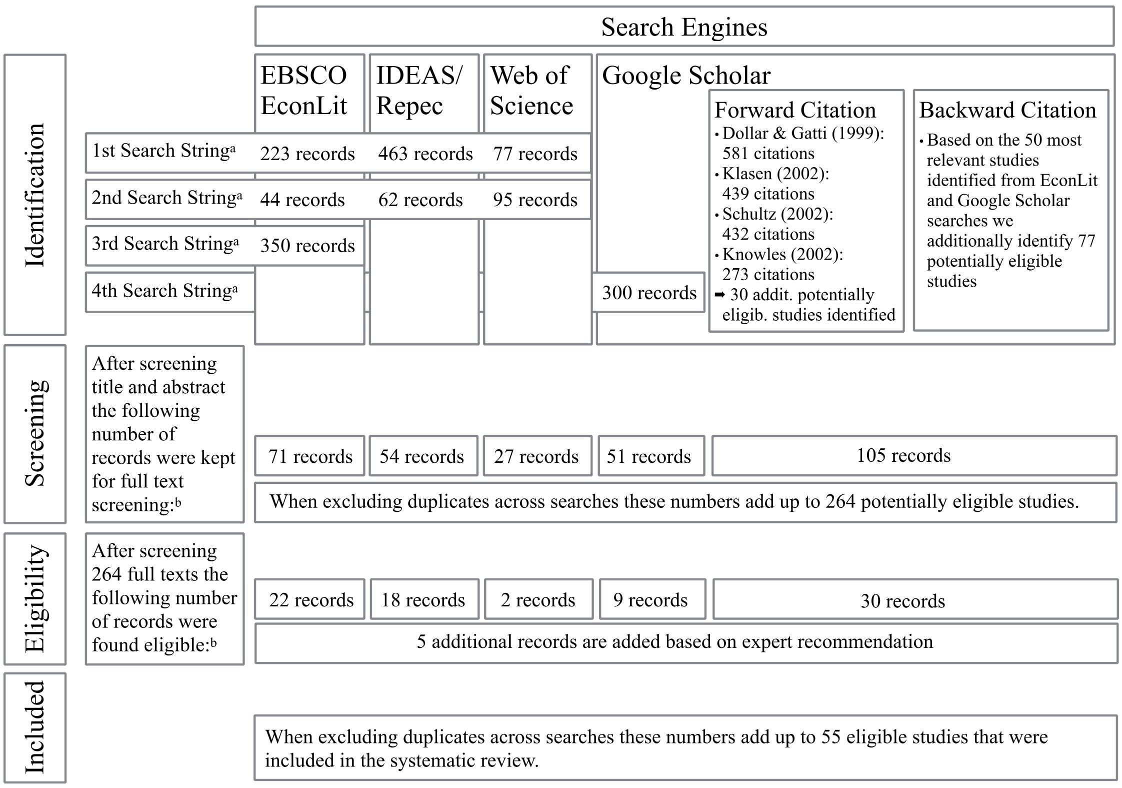
In the first screening, titles and abstracts were screened, only removing those records, which were clearly not relevant for the review based on the criteria above. This led to 308 relevant studies. The removal of duplicates across searches led to a reduction to 264 studies.

Second, 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.. Additionally, reviewers noted when there was uncertainty on how to classify one of the criteria. The two independent ratings were then compared and cases where the eligibility rating differed across reviewers as well as cases classified as unclear were discussed together with a third reviewer (an expert on the topic) for a final inclusion decision. After merging the two reviewer's eligibility assessments and discussing unclear cases among the entire team, and adding 5 records based on expert recommendations, 55 studies published in journals, as working papers, as books, or doctoral theses were decided to be relevant for the synthesis. A large amount of papers was excluded

[8] 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.

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 solely have female or male education (but not both) 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.

Figure 1: Overview of the literature search



[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.

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. Seventeen of the studies use time series methods for single countries, one study uses Bayesian model averaging, three studies run within-country cross-sectional regressions, while the remaining 34 studies cover a larger set of countries using cross-section or panel methods. For comparability reasons, we consider only these 34 cross-country studies for the meta-analysis presented in sections five and six. The time series studies are briefly summarized in section seven.

4. Meta-regression analysis methodology

4.1 Data extraction and sample description

The 34 studies that are eligible for the meta-analysis report a total of 383 regression equations that investigate the educational gender equality and growth relationship. 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 Appendix 2.

The question whether the gender gap in education affects economic growth is assessed in two common ways in our sample. As shown in Table 1, about 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.[9] We will refer to these as the *gap*-studies. As these approaches are fundamentally different, we perform separate analysis for each set of studies, respectively.[10]

The studies included in our sample, further, differ in the choices of how education is measured and which methods are employed. The majority of our studies uses 'quantitative' education measures (e.g. enrollment shares or years of schooling) based on various data sets compiled by Barro and Lee. Only one study in our sample uses literacy information — a measure arguably more focused on education quality. Generally speaking, there is plenty of evidence for the effect of 'quantitative' education gaps on economic growth, while evidence for the effect of 'qualitative' education gaps (e.g. gaps in literacy or math and science test scores) barely exists. Further, 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, 13 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

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

[10] 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.

Estimates, Three Stage Least Squares, Chamberlain´s Pi-matrix, Iteratively Reweighted Least Squares, and Semi-parametric Partially Linear Regression.

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.

4.2 Summarizing effect sizes

In order to summarize the research findings in our sample, we have to find a way to make regression coefficients comparable across regression equations and studies. In observational studies of the kind investigated here, this is usually complicated by the fact that effect sizes are based on regression equations that differ in terms of scales and measures. We therefore 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 and 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 consequently 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 effect 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 underlying effect. 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 of the true underlying average effect (i.e. a constant) as well as an error term:

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

This equation can be further augmented with weights that reflect precision in the estimation. 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 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 suffers from neglecting that observational macroeconomic studies greatly differ, e.g., in terms of sample composition, estimation method, periods, and specification. It is likely that the true underlying effect size varies with these study characteristics. We, therefore, augment our model in (2) by including *random effects* – 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 allow for heterogeneity between studies. We use the Random Effects Maximum Likelihood (REML) estimator, which controls for the between-study variance (Thompson and Sharp 1999, Benos and Zotou 2014, Gallet and Doucouliagos, 2017).[11] 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 and Sharp, 1999, Borenstein et al. 2010, Stanley and 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 errors at study-level.

4.3 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 expectation regarding certain test results on the side of the authors, reviewers, or the editor (Stanley and 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 together 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 as

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

[11] 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.

where r_{ij} is again the partial correlation coefficient and the constant term β_0 again represents a genuine average effect of gender education gap on 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 Figure 1, 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 genuine average effect conditionally on controlling for a possible publication selection, or the precision-effect test (PET). However, Stanley (2008) reports that β_0 in equation (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 equation

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

4.4 Heterogeneity

Furthermore, we augment equation (3) with a vector of moderator variables to explain the heterogeneity in the effect sizes, r_{ij}. We presume that the true underlying effect size varies with characteristics regarding specification (e.g included covariates), regression method (e.g. OLS cross-section regression, fixed effects panel regression) and measurement differences (e.g. type of education variable). We extent equation (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 6.

5. Comparing female and male education coefficients ("*Comparative*" studies)

A number of 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 effect of educational gender inequality 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.

5.1 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 promotes economic growth, and even a smaller minority that it does so more than male education. In fact, male education has a positive and significant impact in more than three times as many studies and in 48 percent of regressions the female education has a significant negative impact on economic growth.

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 puzzling result, that female education is seemingly correlated with economic growth in a negative way while the correlation with male education is positive and statistically significant, was first found in an influential study by Barro and Lee (1994). 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 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 the period average. These features appear to drive the conclusions on the negative effect of female education on growth.[12]

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).While, 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, Lorgelly, and Owen (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 dummy variables would assign the cause of low growth in Latin America to high initial female education there, and conversely high growth in East Asia

[12] In one regression, he includes regional dummies and then the negative effect of female education disappears.

to comparatively low initial female education. Clearly, this is a dubious causal attribution as many factors contributed to the East Asian economic miracle (e.g. World Bank, 1994) and Latin America's poor growth record (e.g. Taylor, 1998), other than initial female education.[13]

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.

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
Logelly 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

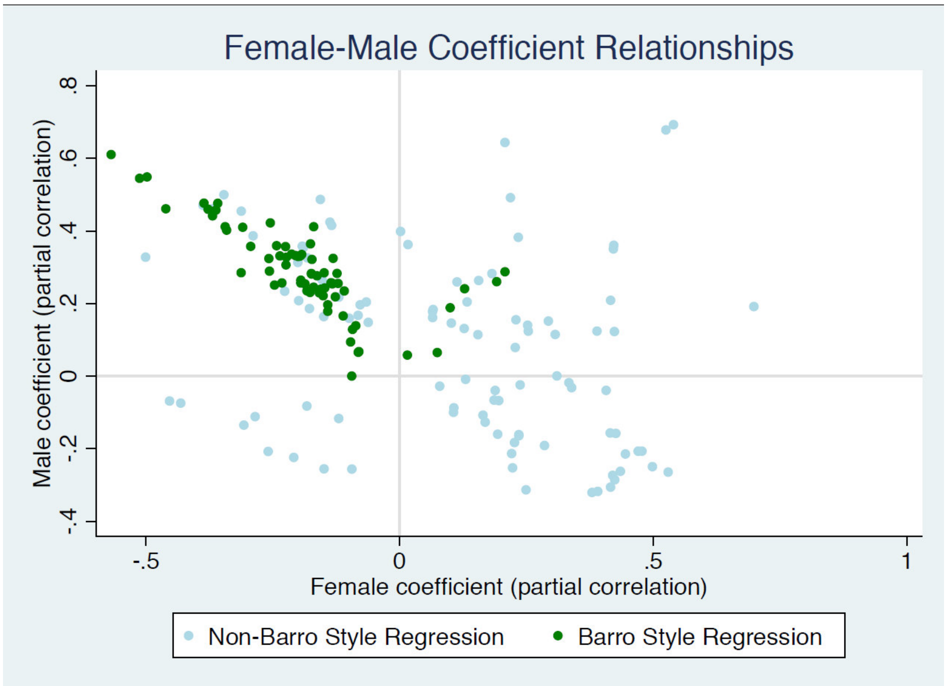
5.2 Graphical analysis

To understand whether our descriptive results in Table 2 are in fact driven by the typical Barro-style specification, we plot coefficient relations that originate from typical Barro-style versus those that origin from other regressions in the sampled studies. To do so, we calculate partial correlations of each of the two education coefficients with the growth variable (as described in equation 1 above) and plot the relationship of the resulting coefficient pairs; see Figure 2 and Figure 3, which show the full set of estimates and the within-study averages of

[13] 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.

estimates, respectively. Similarly 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 affects growth positively while female education affects it 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.[14] Figure 3 shows that also study-average effects using Barro-style specifications are drive most results in the upper left quadrant.

Figure 2: Coefficient relationships, all estimates – Barro-specifications vs. non-Barro-specifications



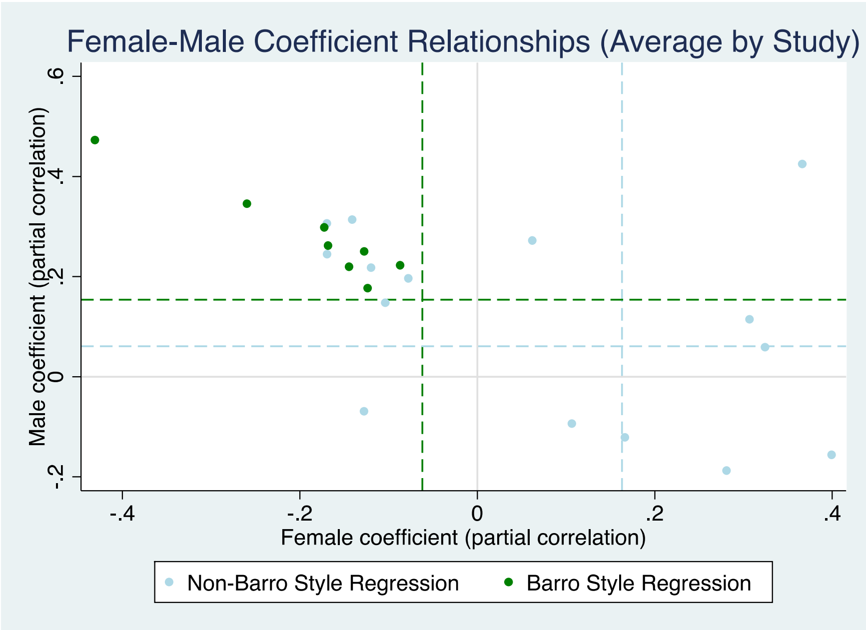
Note: The green and light blue dots show the pair of male and female partial correlation coefficients of education with growth for Barro and non-Barro style regressions, respectively.

The green dashed lines in Figure 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. It becomes evident that the Barro-specifications dominate the plots as they lead to male-positive (PEESE: 0.154; p-value<0.01) versus female-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 coefficient sizes and signs vary notably with

[14] As can be seen in Figure 2, there are a few Barro-style regressions showing a positive correlation 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.

different covariates and methodology. However, when looking at the PEESE estimates 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 has a significant impact on economic growth while male education does not, suggesting that reducing gender inequality in education would boost economic growth.

Figure 3: Coefficient relationships, averaged by study – Barro-specifications vs. non-Barro-specifications



Note: The green and light blue dashed lines in Figure 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.

5.3 Miscellaneous comparative studies

Before turning to the studies using the gender gap in education as covariates in cross-country regressions, 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 impact of education gaps 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 reduces GDP. Another study uses panel fixed effects regressions using annual data for India's states and finds that female literacy leads to significantly higher income in 10 out of 14 specifications while male literacy is never significantly affecting income levels (Esteve-Volart, 2004). A last study for 67 Turkish provinces using 5 year-averages from 1975-2000 show that both female and male education affect GDP positively and significantly, but that only male education has such an effect in less developed provinces, and female education in more developed ones (Tansel and Gungor, 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 1960-1996, 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 has a negative effect on growth (Abington, 2014).[15]

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

A total of seventeen studies, including 216 relevant regressions, present educational inequality measured as a gap variable (e.g. a ratio of female education over male education). Using a gap variable (instead of two separate indicators analyzed in section 5) as measure for the educational gender equality has two advantages: First, it allows for a direct estimate of the impact of educational gender equality on growth. Second, it helps to avoid the problem of multicollinearity, which arises when including, 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).

6.1 Descriptive evidence

Most regression equations investigated here (210 out of 216) include a female-male education *ratio* to measure the gender gap. As the gap-variable is defined as the ratio of female education over male education, an increase in this variable represents an increase in the female relative to male education. Only six regressions use the log *difference* in male and female education, which we manually convert to female over male education. Descriptively, these specifications support the claim that reducing the gender gap in education promotes economic growth: In 80 percent of the cases that uses the female-to-male ratio (*F/M*) of education, the respective coefficient is positive and statistically significant at the conventional level; in only 2.5% of the cases it is negative and significant.[16] Further, in three out of the six estimates that measure equality as a logged difference (log *M* – log *F*), which we converted to the female-over-male coefficient, the effect is positive and statistically significant at the conventional level. A first assessment of the pooled partial correlation coefficient (as described in equation 1) confirms that, by and large, lower inequality may be good for growth: The average partial correlation between the coefficient of the educational gender gap

[15] 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.

[16] 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.

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.

Table 4: *Gap*-studies – Descriptive summary of results

	Share of coefficients…	
Indicator	**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 6. *Significant* refers to statistical significance at least at the 10 percent level (p-value < 0.1).

6.2 Meta-analysis and assessment of publication bias

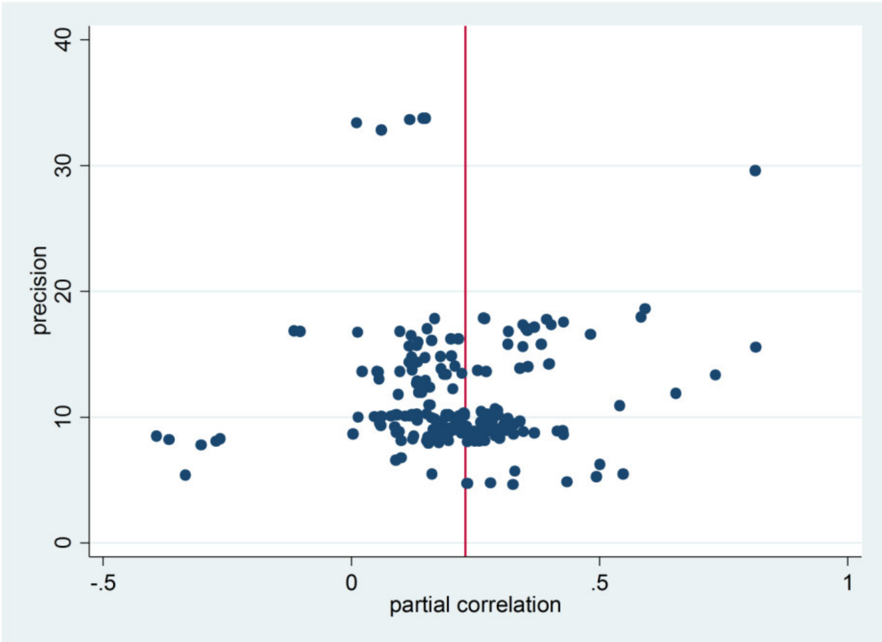
In Table 5 we report the average effects of the educational gender gap with economic growth using several standard meta-analysis techniques, as described in section 4. All models estimate standard errors clustered at the study-level. Column 1 in Table 5 displays the average partial correlation coefficient using a fixed effects meta-analysis model, i.e. a simple OLS estimation without weights, as described in equation (2).[17] 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 of the educational gender-gap with economic growth, ranging from 0.21 to 0.22, which represents a moderate effect (Doucouliagos, 2011).[18] Yet, as described above, it is a reasonable assumption that publication bias and outliers may affect these estimates of the true underlying effect.

Assessing the distribution of our estimates graphically can give a first impression of whether these two concerns are relevant in our sample. The *funnel plot* in Figure 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 and Havránek, 2013; Stanley and 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 effect, while asymmetries may signal publication bias or outliers. The funnel plot in Figure 4 shows that there are no strong asymmetries surrounding the average effect size (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.

[17] 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.

[18] 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.

Figure 4: Funnel plot



Publication bias. To assess publication bias formally, we apply the FAT-PET-PEESE strategy as described above. We report the FAT-PET test for publication bias in column 3 of Table 5. The augment our meta-analysis by including the standard error (*SE*) of the partial correlation coefficient as an explanatory variable. Hence, FAT-PET controls for 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 effect size is robust to this adjustment. However, the coefficient loses statistical significance at conventional levels (p-value = 0.15). As the FAT-PET method tends to underestimate a possible true underlying effect, 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. The estimate that represents the underlying genuine effect of educational gender equality on growth is robust in size and statistically significant at the five percent level.[19]

Outliers. The funnel plot in Figure 4 shows that there are a few estimates that might be outliers in our sample. As the FAT-PET-PEESE test can be affected by outliers, we run the test for publication bias one more time 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. The test results previously described are overall robust to this alteration, i.e. publication bias is not a strong concern in our sample. Yet,

[19] Stanley (2008) notes that if FAT-PET fails to find a genuine average effect 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.

removing outliers does reduce the average effect size, while the coefficients of the publication bias indicators change signs but remains statistically insignificant.

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.

Random effects. Finally, we augment our model by estimating it as random effects model in columns 7 to 10 to control for heterogeneity in our regression coefficients. The specifications discussed so far (columns 1 to 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 and Doucouliagos 2015, 2016). In other words, we assume that there is a single underlying effect size, which is true for all the samples and years of all studies in the meta-analysis. However, this is not necessarily the case as the studies in our sample are very different in terms of included countries, measures, methods and data sets used. The true effect may vary between studies, i.e. the effect size could be higher or lower, depending on whether authors, for instance, compose a data set with a slightly richer or better educated set of countries, or if education and income variables are measured differently, etc. Hence, we estimate our regression using a *Random Effects MRA* to allow for the true effect to vary between studies.. The random effects model assumes that the underlying effects of the seventeen studies included in our MRA are a random sample from a relevant distribution of effect sizes, while the model estimates the mean effects of this distribution (Borenstein et al., 2010). As evident from columns 7 to 10, the results of this analysis are very similar to those previously discussed: We do not find any statistical significance for publication bias, outliers upwardly bias the mean estimate of the underlying true effect 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 summary, we find that the average effect 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 will continue to report conservative estimates controlling for *SE* and compare our results with and without outliers, as well as with and without random effects.

6.3 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. In this section we investigate how these differing characteristics moderate our effect size 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 ourselves to those that we regard as the most relevant ones, and combine some characteristics. Table 6 provides the description of these characteristics.

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

Among other things, Table 6 shows that 27 percent of authors in the included estimates are female, suggesting that the share of female authors in our studies does not differ greatly from the share of female academics in economics.[20] About half of the estimates origin 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 origin from regression equations that use economic variables as covariates (i.e. trade, government expenditure, inflation, macro-economic stability). Only fifty six percent of estimates origin from studies that include control variables for social and institutional variables such as democracy, rule of law, human/women rights, religion and the like. Sixty three percent of

[20] 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)

regressions measure education in terms of school enrollment, including primary, secondary and tertiary schools, while 37 percent use other measures of education such as schooling attainment or literacy rates.

We explore heterogeneity in effect sizes 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 effects, such as publication status and 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 average effect size (the constant term) decreases when estimates origin 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 leads to a smaller effect size. Inclusion of variables for female labor force participation (FLFP), initial education, and social and institutional controls increase the effect size. These results are robust to removing outliers, except for the coefficient on FLFP, which no longer seems to have a moderating effect (see column 4). Turning to the random effects model in 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 effect, statistically significant at the one percent level. Overall, removing outliers yields similar results, while coefficients are generally smaller (column 5).

In Table 8, we perform similar analyses with the second set of control variables related to measurement. Generally, these regressions perform worse compared to Table 7 in terms of goodness of fit and significant effects. In columns 1 and 2, we find that in the fixed effects and random effects MRA, the average effect size is larger when estimates origin from regression equations where education is measured in terms of enrollment (i.e. versus alternative measures like schooling attainment). Other measurement choices do not seem to have any influence on the effect size. These conclusions are generally robust to removing outliers in column 4 and 5.

Table 7: Effect size estimates, including specification-related moderators

	(1)	(2)	(3)	(4)	(5)
				Without outliers	
	Partial correlation coefficient				
Constant	0.207**	0.272***	0.272**	0.158***	0.227***
	(0.071)	(0.042)	(0.094)	(0.045)	(0.077)
SE	-0.662	-0.651*	-0.651	0.072	-0.184
	(1.013)	(0.365)	(0.782)	(0.671)	(0.635)
FLFP	0.091**	0.043	0.043	0.029	0.027
	(0.042)	(0.035)	(0.042)	(0.034)	(0.033)
Fixed effects	-0.099**	-0.079***	-0.079*	-0.068*	-0.065*
	(0.041)	(0.030)	(0.041)	(0.035)	(0.035)
Share of female authors	-0.162*	-0.164***	-0.164**	-0.135*	-0.140**
	(0.081)	(0.039)	(0.071)	(0.068)	(0.059)
Published	-0.016	-0.025	-0.025	-0.019	-0.028
	(0.041)	(0.031)	(0.040)	(0.037)	(0.035)
Economic controls	-0.072	-0.134***	-0.134**	-0.075	-0.119**
	(0.076)	(0.034)	(0.058)	(0.053)	(0.046)
Initial education	0.228***	0.228***	0.228***	0.200***	0.202***
	(0.060)	(0.035)	(0.070)	(0.053)	(0.058)
Social/Institutional controls	0.114***	0.098***	0.098**	0.102***	0.091***
	(0.034)	(0.026)	(0.034)	(0.028)	(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 are clustered at the study level. In c4 and 5 we exclude outliers defined as those estimates which lie beyond an absolute standard deviation larger than 3.5.

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

	(1)	(2)	(3)	(4)	(5)
				Without outliers	
	Partial correlation coefficient				
Constant	0.184	0.152***	0.152	0.097	0.094
	(0.115)	(0.054)	(0.111)	(0.070)	(0.081)
SE	-0.564	-0.078	-0.086	0.404	0.486
	(0.989)	(0.418)	(0.971)	(0.577)	(0.703)
Enrollment	0.117**	0.097***	0.097***	0.088***	0.083***
	(0.042)	(0.030)	(0.033)	(0.030)	(0.027)
Dep. var.: Levels	0.056	0.029	0.030	0.061	0.049
	(0.067)	(0.027)	(0.076)	(0.052)	(0.058)
Dep. var.: Logs	-0.061	-0.020	-0.021	-0.042	-0.024
	(0.042)	(0.024)	(0.044)	(0.035)	(0.036)
Source: Barro	0.009	0.004	0.005	0.028	0.021
	(0.074)	(0.035)	(0.065)	(0.052)	(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 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.

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

Alternative effect size calculation. So far, we have estimated the true underlying effect 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 clse to normally distributed; but for robustness we report results on the transformed effect size as well. The results of our main analysis based on this transformation are presented in Table 9. While the coefficients in all presented tests are decreasing compared to the results in Table 5, partly reducing the effect size from moderate to small, our qualitative conclusions from above still hold.

Table 9: 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 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.

Full set of moderators. In Table 10 we further test robustness of the heterogeneity analysis once all covariates are included, to see which of the moderators that had significant effects in 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 there are changes regarding the measurement-related moderators of Table 8. Specifically, in column 1 covariates that increase the size of the effect of gender education gap on growth are *Social/Institutional controls* and *Initial education* as before, statistically significant at the one percent level while the effect of *Enrollment* is not statistically significant anymore. The variables that decrease the effect size are *Fixed effects*, *Share of female authors* and *Economic controls* in both fixed effects and random effects regressions.

Columns 3 and 4 of Table 10 show the results for the Z-transformed dependent variable. As one can observe from the table the coefficient of the average effect (the constant term) is further reduced as a result of the Z-transformation. The weighted OLS fixed effects results are similar of that in column 1, with slight differences in statistical significance for some variables. The results of the Z-transformed random effects regression show some differences compared to those in column 2. In column 4, the source of data – Barro and Lee (1993, 1996, 2001, 2013) – is statistically significant while in the specification that uses the non-transformed effect sizes it is not.

Table 10: Effect size estimates, including all moderators

	(1)	(2)	(3)	(4)
	Without outliers			
	Partial correlation		**Z-transformed**	
Constant	0.163**	0.153**	0.137**	0.112**
	(0.057)	(0.064)	(0.061)	(0.046)
SE	-0.101	0.162	0.211	0.542
	(0.515)	(0.553)	(0.552)	(0.357)
FLFP	0.008	0.002	0.023	0.017
	(0.023)	(0.027)	(0.023)	(0.030)
Fixed effects	-0.059***	-0.055***	-0.049***	-0.039
	(0.016)	(0.019)	(0.015)	(0.025)
Share of female authors	-0.131*	-0.113*	-0.131**	-0.114***
	(0.063)	(0.059)	(0.061)	(0.038)
Published	0.010	0.023	0.024	0.043
	(0.055)	(0.061)	(0.064)	(0.034)
Economic controls	-0.097**	-0.101**	-0.104***	-0.099***
	(0.039)	(0.036)	(0.036)	(0.031)
Initial education	0.232***	0.212***	0.249***	0.233***
	(0.062)	(0.063)	(0.064)	(0.033)
Social/Institutional controls	0.074**	0.057**	0.054*	0.047*
	(0.027)	(0.027)	(0.027)	(0.025)
Enrollment	0.049	0.042	0.048	0.041
	(0.029)	(0.031)	(0.034)	(0.027)
Dep. var.: Levels	-0.025	-0.017	-0.040	-0.041
	(0.054)	(0.062)	(0.068)	(0.032)
Dep. var.: Logs	0.062	0.057	0.098*	0.093***
	(0.039)	(0.043)	(0.055)	(0.026)
Source: Barro	-0.092	-0.092	-0.121	-0.134***
	(0.065)	(0.073)	(0.080)	(0.039)
Weights	Yes	Yes	Yes	Yes
Random effects		Yes		Yes
Small cluster adj.		Yes		Yes
No. of studies	17	17	17	17
(Adj.) R^2	0.474	0.817	0.482	0.576
Observations	212	212	212	212

Notes: Standard errors 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.

Based on these tests for robustness we can conclude that, on average, there is a statistically significant correlation between reducing gender gap in education and economic growth. The size of the effect is increased 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 effect 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 features, they will have a larger than average partial correlation coefficient, i.e. they will increase the effect of gender gaps on growth.

It is also notable that studies with a high female share in authorship find smaller effects of the educational gender gap on growth which is sizable and robust across all specifications. We tested whether this is due to a particular female author and found this not to be the case. There is a sizable literature on gender gaps in economics, including gender differences in the publication process (e.g. Hengel, 2017 and studies cited therein), and there is work suggesting that the sex of the experimenter affects results 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.

7. 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), Chaudry (2007), Qureshi et al. (2007) use ratios of female to male education indicators as explanatory variables, mostly enrollment ratios and in the case of Chaudry (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. Chaudry (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 to 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 (2012a) 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 had a causal impact on 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 causal impact on 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 has a growth promoting effect. We decided not to include the time series studies in the meta-analysis because of comparability issues, because of a high heterogeneity in methods and 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.

8. Concluding remarks

The literature that specifically focuses on the impact of the gender gap in education on growth is small but growing, and relatively recent. It is largely based on cross-country regressions and thus can suffer 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 impacts for 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. The heterogeneity of studies using different methods and data makes it important to discern whether one can identify robust effects.

This paper presents a systematic review and a meta-analysis of the this literature. Although many policymakers and international organizations by now accept as a fact that gender inequality in education can hurt economic growth, the research findings are not always conclusive. In addition, research findings also suffer from a publication bias. We find 55 studies that include a regression analysis of gender differences in education and economic output. However, these studies are very heterogeneous in their samples, methods, measurements and econometric models. Therefore, we first split the sample into two subgroups: single country studies and cross country studies. In the case of single country studies, we provide a brief review of the findings and argue that they only provide a very partial and unreliable picture of effects. We restrict our meta- analysis to the cross-country samples. Yet, within this sample a group of studies explores the link between the educational gender gap and economic growth and the other group of studies explores the separate effects of female education and male education on economic growth. Since the purpose of this study is to look at the evidence on the link between gender inequality in education and growth, we run a meta-regression analysis for the first group of studies. We also conduct a separate analysis for the group of studies that look at the separate effects of female education and male education on growth. There we find that the average effect of female and male education on growth depends greatly on whether a Barro-style specification was used or not. We argue that a Barro-style specification is likely to suffer from left-out variable bias, esp. related to regional differences in economic performance. When excluding these studies, there is a positive and significant average effect of female education on growth, while the smaller positive effect of male education is not significant.

For the purpose of meta-analysis we collect data from all the regression analysis in the 17 papers that use the female-male ratio of education and end up with 216 effect size estimates in total. We estimate the effect size using fixed effect and random effect MRA estimation and we cluster the standard errors at the study level. We also estimate the effects of moderating variables. In result, we find that there is a positive correlation between gender equality in education and growth, yet the size of the effect is smaller if more robust methods are used, such as panel and time fixed effects. However, it becomes larger once non-economic but relevant variables are included in model, such as the institutional environment or initial education levels. We find a persistent negative effect of the share of female authors on the effect sizes that deserves further scrutiny in future research.

References

Barro and 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, Robert and Jong-Wha Lee, 2013, "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*104:184-198

Barro, Robert and Jong-Wha Lee. 1994. Sources of Economic Growth." *Carnegie-Rochester Conference Series on Public Policy*, 40; 1-46

Barro, Robert J., and Jong Wha Lee. 1996. "International measures of schooling years and schooling quality." *The American Economic Review* 86(2): 218-223.

Barro, Robert J., and Jong Wha Lee. 2013. "A new data set of educational attainment in the world, 1950–2010." *Journal of Development Economics* 104: 184-198.

Barro, Robert J., and Jong-Wha Lee. 1993. "International comparisons of educational attainment." *Journal of monetary economics* 32(3): 363-394.

Barro, Robert J., and Jong-Wha Lee. 2001. "International data on educational attainment: updates and implications." *oxford Economic papers* 53(3): 541-563.

Becker G. (1981) "*A treatise on the family*." Chicago: University of Chicago Press.

Benos, Nikos, and Stefania Zotou. 2014. "Education and economic growth: A meta-regression analysis." *World Development* 64: 669-689.

Bloom, David and Jeffrey Williamson. 1998. "Demographic Transition and Economic Miracles in Emerging Asia." *The World Bank Economic Review*12(3): 419-455.

Borenstein, Michael, et al. 2010. "A basic introduction to fixed-effect and random-effects models for meta-analysis." *Research synthesis methods* 1(2): 97-111.

Brockwell, Sarah E., and Ian R. Gordon. 2001. "A comparison of statistical methods for meta-analysis." *Statistics in medicine* 20(6): 825-840.

Busse, Matthias and Christian Spielmann. 2006. "Gender inequality and trade." *Review of International Economics* 14(3): 362-279.

Chapman, C., C. Benedict, and H. Schiöth. 2018. Experimenter gender and replicability in science. *Science Advances* 4(1), e1701427.

Doucouliagos, C. 2011. "How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics." Economics Series Working Paper 05, Deakin University.

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, Matthias, et al. 1997. "Bias in meta-analysis detected by a simple, graphical test." *Bmj* 315(7109): 629-634.

Gallet, Craig A., and Hristos Doucouliagos. 2017. "The impact of healthcare spending on health outcomes: A meta-regression analysis." *Social Science & Medicine* 179: 9-17.

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.

Iršová, Zuzana, and Tomáš Havránek. 2013. "Determinants of horizontal spillovers from FDI: Evidence from a large meta-analysis." *World Development* 42: 1-15.

King, E. S. Klasen, and M. Porter. 2009. *Women and Development.* In Lomborg, B. (ed.) Global Crises,m Global Solutions, 2nd edition. Cambridge: Cambridge University Press.

Klasen, S. (2006) "Gender and Pro-Poor Growth." In Menkoff, L. (ed.) Pro-Poor Growth: Evidence and Policies. Berlin: Dunker&Humblot.
Klasen, S. 2016. "Gender, Institutions, and Economic Development." *Courant Research Center Discussion Paper.*
Lagerlöf, N. 2003. ''Gender Equality and Long-Run Growth.'' *Journal of Economic Growth* 8(4): 403–26.
Levine, R., & Renelt, D. (1992). "A sensitivity analysis of cross-country growth regressions." *The American economic review*,:942-963.
Lorgelly, P. and D. Owen. 1999. "The Effect of Female and Male Schooling on Economic Growth in the Barro-Lee Model." *Empirical Economics*, 24, August.
Murthi, Mamta, Anne-Catherine Guio, and Jean Drèze. 1995. "Mortality, Fertility, and Gender Bias in India: A District-Level Analysis." *Population and Development Review* 21(4): 745-782.
Petticrew, M. and Roberts, H. *Systematic reviews in the social sciences: a practical guide*.Oxford. Blackwell Publishing. 2006.
Romero, J. 2013. "Where Are the Women?" *Econ Focus* 7 (2),: 12.
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.
Schultz, T. P. (2002). "Why governments should invest more to educate girls." *World development*, 30(2): 207-225.
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.
Seguino, S. 2000a. "Gender Inequality and Economic Growth: A Cross-Country Analysis." *World Development* 28: 1211-1230.
Seguino, Stephanie. 2000b. "Accounting for Gender in Asian Economic Growth*", Feminist Economics, Taylow and Francis Journals*6(3):27-58.
Stanley, Tom D. 2005. "Beyond publication bias." *Journal of economic surveys* 19(3): 309-345.
Stanley, Tom D. 2008. "Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection." *Oxford Bullettin of Economics and Statistics*, 70 (1):103-127
Stanley, Tom D. 2017. "Limitations of PET-PEESE and Other Meta-Analysis Methods. " *Social Psychological and Personality Science*, 8 (5) : 581-591
Stanley, Tom D., Doucouliagos, Hristos. 2012. "*Meta-regression Analysis in Economics and Business*." Routledge, Oxford (2012)
Stanley, Tom D., Doucouliagos, Hristos. 2015. "Neither fixed nor random: weighted least squares meta-analysis. " *Statistics Med.*, 34 (13) :2116-2127
Stanley, Tom D., Doucouliagos, Hristos. 2016. " Neither fixed nor random: weighted least squares meta regression analysis. " Res. Synthesis Methods. forthcoming
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
Teignier, M. and D. Cuberes. 2015. "Aggregate costs of gender gaps in the labor market: A quantitative excercise." *Journal of Human Capital* (forthcoming).
Tertilt, M. and M. Doepke. 2014. "Does female empowerment promote economic development?" NBER Working Paper No. 19888.
Thompson, Simon G., and Stephen J. Sharp. 1999. "Explaining heterogeneity in meta-analysis: a comparison of methods." *Statistics in medicine* 18(20): 2693-2708.

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.

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 (1996b) 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 Resaerch 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. (2000). Accounting for Gender in Asian Economic Growth. *Feminist Economics*, *6*(3), 27–58.

Seguino, S. (2000). 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*. University of California, Davis.

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.

APPENDIX 1 – Search strategies

1. EconLit

1st Search String (Conventional)

Tuesday, April 05, 2016 7:42:57 AM

#	Query	Limiters/Expanders	Last Run Via	Results
S9	(S1 OR S2) AND (S3 OR S4) AND (S5 OR S6) AND (S7 OR S8)	Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	223
S8	AB educat* OR AB school* OR AB learn* OR AB universit* OR AB college* OR AB enrol* OR AB ("human capital") OR AB literate OR AB literacy OR AB attain* OR AB student* OR AB pupil*	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	114,538
S7	TI educat* OR TI school* OR TI learn* OR TI universit* OR TI college* OR TI enrol* OR TI ("human capital") OR TI literate OR TI literacy OR TI attain* OR TI student* OR TI pupil*	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	47,871
S6	AB gap OR AB equity OR AB equal* OR AB unequal* OR AB inequal* OR AB ratio* OR AB discriminat* OR AB differential* OR AB ("female-to-male") OR AB ("female-male") OR AB ("male-to-female") OR AB	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	79,211

	("male-female")			
S5	TI gap OR TI equity OR TI equal* OR TI unequal* OR TI inequal* OR TI ratio* OR TI discriminat* OR TI differential* OR TI ("female-to-male") OR TI ("female-male") OR TI ("male-to-female") OR TI ("male-female")	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	24,743
S4	AB girl* OR AB female* OR AB women* OR AB gender* OR AB sex*	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	24,629
S3	TI girl* OR TI female* OR TI women* OR TI gender* OR TI sex*	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	10,963
S2	AB ("economic growth") OR AB ("economic development") OR AB ("economically grow") OR AB ("economically develop") OR AB ("economic performance") OR AB GDP OR AB GNP OR AB GNI OR AB ("gross domestic product") OR AB ("gross national income") OR AB ("gross national product")	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	53,265
S1	TI ("economic growth") OR TI ("economic development") OR TI ("economically grow") OR TI ("economically	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	57,246

	develop") OR TI GDP OR TI GNP OR TI GNI OR TI ("gross domestic product") OR TI ("gross national income") OR TI ("gross national product") OR TI ("economic performance") OR TI growth			

2nd Search String (Conventional)

Monday, June 26, 2016 1:15:57 AM

#	Query	Limiters/Expanders	Last Run Via	Results
S4	S1 AND S2 AND S3	Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	44
S3	TI ("economic growth") OR TI ("economic development") OR TI ("economically grow") OR TI ("economically develop") OR TI GDP OR TI GNP OR TI GNI OR TI ("gross domestic product") OR TI ("gross national income") OR TI ("gross national product") OR TI ("economic performance") OR TI growth	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	57,963
S2	TI girl* OR TI female* OR TI women* OR TI gender* OR TI sex*	Expanders - Apply equivalent subjects; Apply related words Search modes - Boolean/Phrase	Interface - EBSCOhost Research Databases Search Screen - Advanced Search Database - EconLit	19,476
S1	TI educat* OR TI school* OR TI learn* OR TI universit* OR TI	Expanders - Apply equivalent subjects; Apply related words	Interface - EBSCOhost Research Databases Search Screen - Advanced Search	48,816

	college* OR TI enrol* OR TI ("human capital") OR TI literate OR TI literacy OR TI attain* OR TI student* OR TI pupil*	Search modes - Boolean/Phrase	Database - EconLit

3rd Search String (Expert-based)

Note that the following table represents a trial of several search strings based on expert considerations. We settled on using the search string S20 for our review.

Tuesday, February 23, 2016 9:12:38 AM

#	Query	Limiters/Expanders	Results
S1	gender	Limiters - Published Date: 20080101-20101231 Search modes - Boolean/Phrase	4,853
S2	gender AND inequality	Limiters - Published Date: 20080101-20101231 Search modes - Boolean/Phrase	420
S3	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth	Limiters - Published Date: 20080101-20101231 Search modes - Boolean/Phrase	157
S4	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth	Limiters - Published Date: 20080101-20101231 Search modes - Boolean/Phrase	157
S5	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth	Limiters - Published Date: 20080101-20091231 Search modes - Boolean/Phrase	91
S6	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth AND AB panel OR AB cross-country	Limiters - Published Date: 20080101-20091231 Search modes - Boolean/Phrase	847
S7	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth AND AB panel OR AB cross-country	Limiters - Published Date: 19600101-19901231 Search modes - Boolean/Phrase	65
S8	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth AND AB panel	Search modes - Find all my search terms	6,689

	OR AB cross-country		
S9	AB gender AND AB *equality OR AB gap AND AB education* AND AB economic growth AND AB panel OR AB cross-country	Limiters - Published Date: 20080101-20091231 Search modes - Find all my search terms	847
S10	AB gender inequality AND AB *equality OR AB gap AND AB education* AND AB economic growth AND AB panel OR AB cross-country	Limiters - Published Date: 20080101-20091231 Search modes - Find all my search terms	791
S11	AB gender inequality OR AB gap AND AB education* AND AB economic growth AND AB panel OR AB cross-country	Limiters - Published Date: 20080101-20091231 Search modes - Find all my search terms	911
S12	AB gender inequality education growth OR AB gender gap education growth AND AB female education* growth	Limiters - Published Date: 20080101-20091231 Search modes - Find all my search terms	9
S13	AB gender* *equality education* growth* OR AB gender* gap* education* growth* AND AB female* education* growth*	Limiters - Published Date: 20080101-20091231 Search modes - Find all my search terms	4
S14	AB gender* *equality education* growth* OR AB gender* gap* education* growth* AND AB female* education* growth*	Limiters - Published Date: 19900101-20141231 Search modes - Find all my search terms	46
S15	AB gender inequality education growth OR AB gender gap education growth AND AB female education* growth	Limiters - Geographic Descriptor: Global Search modes - Find all my search terms	1
S16	AB education* *equality gap* gender* growth*	Limiters - Published Date: 20080101-20091231 Search modes - Find all my search terms	1
S17	AB education* *equality gap* gender* growth*	Search modes - Find all my search terms	9
S18	AB education* *equality gender* growth* OR AB education* gap* gender* growth* OR AB education* female* growth*	Search modes - Find all my search terms	306
S19	AB education* *equality gender* growth* OR AB education* gap* gender* growth* OR AB education* female* growth* OR AB school* female* growth*	Search modes - Find all my search terms	357

S20	AB education* *equality gender* growth* OR AB education* gap* gender* growth* OR AB education* female* growth* OR AB school* female* economic growth* OR AB school* girl* economic growth	Search modes - Find all my search terms	350
S21	AB education* *equality gender* growth* OR AB educat* gap* gender* growth* OR AB educat* female* growth* OR AB school* female* economic growth* OR AB school* girl* economic growth OR AB learning* differ* growth* gender* OR AB learn* gap* income women	Search modes - Find all my search terms	365
S22	AB education* *equality gender* growth* OR AB education* gap* gender* growth* OR AB education* female* growth* OR AB school* female* economic growth* OR AB school* girl* economic growth OR AB learni* differ* growth* gender* OR AB learn* gap* income women OR AB educ* *equality women GDP	Search modes - Find all my search terms	357

2. IDEAS

1st Search String

(gender | female | girl | women | sex)

+ (gap | equity | equality | inequality | equal | unequal | ratio | differential | discriminate | discrimination | "female-to-male" | "female-male" | "male-to-female" | "male-female").

+ (education | educating | schooling | school | learn | literate | literacy | university | college | enroll | enrollment | "human capital" | attain | attainment | student | pupil)

+ (GDP | GNI | GNP | "economic growth" | "economic performance" | "economic development" | "gross domestic product" | "gross national income" | "gross national product")

Search results: gender : 35164, gendering : 72, gendered : 1304, genders : 812, gender's : 0, female : 21188, females : 8764, femaleness : 10, female's : 0, girl : 296, girls : 4160, girl's : 0, women : 62170, womens : 86, women's : 0, sex : 9408, sexed : 19, sexes : 949, gap : 37260, gaping : 18, gaped : 1, gaper : 1, gaps : 10067, gap's : 0, equity : 44372, inequity : 1445, equities : 1952, equality : 7167, equalities : 171, equality's : 0, inequality : 52434, inequalities : 9456, equal : 19574, equally : 7711, equaling : 42, equaled : 62, equals : 2749, unequal : 3734, unequally : 539, unequaled : 6, ratio : 41057, ratios : 15596, ratio's : 0, differential : 19208, differentially : 920, differentials : 9563, differential's : 0, discriminate : 2338, discriminative : 136, discrimination : 13868, discriminations : 174, discriminating : 1064, discriminated : 489, discriminates : 267, discrimination : 13868, discriminate : 2338, discriminative : 136, discriminations : 174, discriminating : 1064, discriminated : 489, discriminates : 267, female : 21188, females : 8764, femaleness : 10, female's : 0, to : stopword, male : 15134, males : 9724, maleness : 13, male's : 0, female : 21188, females : 8764, femaleness : 10, female's : 0, male : 15134, males : 9724, maleness : 13, male's : 0, male : 15134, males : 9724, maleness : 13, male's : 0, to : stopword, female : 21188, females : 8764, femaleness : 10, female's : 0, male : 15134, males : 9724, maleness : 13, male's : 0, female : 21188, females : 8764, femaleness : 10, female's : 0, education : 95788, educate : 559, educative : 120, educations : 153, educating : 537, educated : 9215, educates : 20, education's : 0, educating : 537, educate : 559, educative : 120, education : 95788, educations : 153, educated : 9215, educates : 20, schooling : 13074, school : 48293, schooled : 62, schooler : 4, schoolers : 21, schools : 18286, school : 48293, schooling : 13074, schooled : 62, schooler : 4, schoolers : 21, schools : 18286, learn : 7711, learning : 43815, learnings : 39, learned : 3836, learner : 343, learners : 841, learns : 510, literate : 390, literation : 1, literately : 0, literateness : 0, literacy : 4794, university : 64016, universities : 13199, university's : 0, college : 13496, colleges : 2246, college's : 0, enroll : 667, enrolling : 285, enrolled : 1906, enrolls : 16, enrollment : 4532, enrollments : 396, enrollment's : 0, human : 72766, humanly : 17, humans : 1843, humanness : 3, capital : 198087, capitally : 1, capitals : 878, attain : 2627, attaining : 1184, attained : 2461, attainer : 0, attainers : 2, attains : 464, attainment : 7870, attainments : 556, attainment's : 0, student : 14842, students : 39551, student's : 0, pupil : 1089, pupils : 1614, pupil's : 0, gdp : 46059, gni : 201, gnp : 2211, economic : 418467, economics : 61288, growth : 277528, grow : 7934, growly : 0, growing : 35085, grower : 514, growers : 2268, grows : 2871, ingrowth : 1, economic : 418467, economics : 61288, performance : 168477, performances : 8438, performance's : 0, economic : 418467, economics : 61288, development : 246408, developments : 25509, development's : 0, gross : 16456, grossly : 287, grossing : 29, grossed : 10, grossest : 3, grosser : 46, grosses : 31, grossness : 2, domestic : 69607, product : 71586, productive : 19381, products : 59944, product's : 0, gross : 16456, grossly : 287, grossing : 29, grossed : 10, grossest : 3, grosser : 46, grosses : 31, grossness : 2, national : 105042, nationally : 3878, nationals : 571, income : 198942, incoming : 1094, incomer : 5, incomers : 24, incomes : 15368, gross : 16456, grossly : 287, grossing : 29, grossed : 10, grossest : 3, grosser : 46, grosses : 31, grossness : 2, national : 105042, nationally : 3878, nationals : 571, product : 71586, productive : 19381, products : 59944, product's : 0.

In: Abstract
From: Any year
To: Any year
Match: Boolean
Synonyms: Yes
Search Time: 2:37pm , April 5, 2016
Search results total: 463

2nd Search String

(gender | female | girl | women | sex)

+(education | educating | schooling | school | learn | literate | literacy | university | college | enroll | enrollment | "human capital" | attain | attainment | student | pupil)

+(growth| GDP | GNI | GNP | "economic growth" | "economic performance" | "economic development" | "gross domestic product" | "gross national income" | "gross national product").

Search results: gender : 9590, gendering : 51, gendered : 304, genders : 21, gender's : 0, female : 3029, females : 199, femaleness : 2, female's : 0, girl : 78, girls : 421, girl's : 0, women : 8306, womens : 13, women's : 0, sex : 1762, sexed : 1, sexes : 72, education : 18844, educate : 39, educative : 16, educations : 7, educating : 129, educated : 360, educates : 2, education's : 0, educating : 129, educate : 39, educative : 16, education : 18844, educations : 7, educated : 360, educates : 2, schooling : 2397, school : 8298, schooled : 6, schooler : 1, schoolers : 4, schools : 2611, school : 8298, schooling : 2397, schooled : 6, schooler : 1, schoolers : 4, schools : 2611, learn : 1311, learning : 11983, learnings : 22, learned : 1186, learner : 53, learners : 101, learns : 16, literate : 13, literation : 0, literately : 0, literateness : 0, literacy : 986, university : 18992, universities : 1960, university's : 0, college : 2626, colleges : 419, college's : 0, enroll : 18, enrolling : 13, enrolled : 31, enrolls : 3, enrol : 4, enrollment : 670, enrollments : 44, enrollment's : 0, human : 13163, humanly : 0, humans : 104, humanness : 1, capital : 36874, capitally : 0, capitals : 90, attain : 36, attaining : 50, attained : 12, attainer : 0, attainers : 0, attains : 1, attainment : 1094, attainments : 47, attainment's : 0, student : 2869, students : 3421, student's : 0, pupil : 94, pupils : 104, pupil's : 0, growth : 51774, grow : 328, growly : 0, growing : 1995, grower : 85, growers : 313, grows : 39, ingrowth : 1, gdp : 2175, gni : 1, gnp : 241, economic : 87709, economics : 24159, growth : 51774, grow : 328, growly : 0, growing : 1995, grower : 85, growers : 313, grows : 39, ingrowth : 1, economic : 87709, economics : 24159, performance : 35237, performances : 999, performance's : 0, economic : 87709, economics : 24159, development : 53313, developments : 4343, development's : 0, gross : 929, grossly : 2, grossing : 5, grossed : 2, grossest : 0, grosser : 3, grosses : 1, grossness : 1, domestic : 5893, product : 10102, productive : 1509, products : 5623, product's : 0, gross : 929, grossly : 2, grossing : 5, grossed : 2, grossest : 0, grosser : 3, grosses : 1, grossness : 1, national : 14654, nationally : 98, nationals : 37, income : 26692, incoming : 32, incomer : 1, incomers : 1, incomes : 1414, gross : 929, grossly : 2, grossing : 5, grossed : 2, grossest : 0, grosser : 3, grosses : 1, grossness : 1, national : 14654, nationally : 98, nationals : 37, product : 10102, productive : 1509, products : 5623, product's : 0.

In: Title
From: Any year
To: Any year
Match: Boolean
Synonyms: Yes
Search Time: 1:20pm , June 26, 2016
Search results total: 62

3. Web of Science – Core Collection

1st Search String

# 4	95	#3 AND #2 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 3	653,281	TI =("economic growth" OR "economic development" OR "economically grow" OR "economically develop" OR GDP OR GNP OR GNI OR "gross domestic product" OR "gross national income" OR TI "gross national product" OR "economic performance" OR growth) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 2	672,336	TI=(girl* OR female* OR women* OR gender* OR sex*) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 1	768,756	TI=(educat* OR school* OR learn* OR universit* OR college* OR enrol* OR ("human capital") OR literate OR literacy OR attain* OR student* OR pupil*) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*

2nd Search String

# 9		#8 OR #7 OR #6 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 8	36	#4 AND #3 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 7	21	#4 AND #2 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 6	40	#3 AND #2 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 5	10	#4 AND #3 AND #2 AND #1 *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 4	759,052	TI=(educat* OR school* OR learn* OR universit* OR college* OR enrol* OR "human capital" OR literate OR literacy OR attain* OR student* OR pupil*) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 3	587,865	TI=(gap OR equity OR equal* OR unequal* OR inequal* OR ratio* OR discriminat* OR differential* OR "female-to-male" OR "female-male" OR "male-to-female" OR "male-female“) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 2	662,983	TI=(girl* OR female* OR women* OR gender* OR sex*) *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
# 1	23,163	TI= ("economic growth" OR "economic development" OR "economically grow" OR "economically develop" OR GDP OR GNP OR GNI OR "gross domestic product" OR "gross national income" OR "gross national product" OR "economic performance") *Indexes=SCI-EXPANDED, SSCI Timespan=All years*

4. Google Scholar

Search String: gender inequality gap education economic growth cross-country analysis
Where: anywhere in text
When: no restrictions
Results: about 26.500 studies

Based on relevance checks we included the first 300 records to enter the screening stage.

APPENDIX 2 –Data extraction

IDs	
Study ID	
Method ID	
Regression ID	
Coefficient ID	
Bibliographic Information	
Authors	
Year	
Title	
Journal/Book/Institution	
Volume	
Issue	
Publication Type	Journal Article, Working Paper, Book Chapter, Conference Proceeding, Dissertation.
Effect Sizes	
Coefficient	
Significance Level	10%, 5%, 1%, not significant.
Standard Error	
Additional Statistic - Type	If standard error is not reported.
Additional Statistic - Value	If standard error is not reported.
Covariates	All covariates included in the regression equation.
Dependent Variable	
Long Description	Entire information provided on the dependent variable in text.
Short Description	As found in table.
Data Source	E.g. the Penn World Tables PWT;, World Bank, World Development Indicators WDI; or a data set which some particular researchers came up with.
Is variable averaged over several years?	Number of years the variable is averaged for, e.g. 5-year averages.
Logarithmic transformation	E.g. log GDP, ln GDP; (Yes=1, No=0)

Is variable in levels?	E.g. level of GDP, GNP, etc; (Yes=1, No=0)
Is variable in growth rates?	E.g. growth rate of GDP, GNP, etc; (Yes=1, No=0)
Currency	E.g. name of local currency, or US dollar 2005, or PPP current international $, etc.
In per capita terms?	(Yes=1, No=0)
Explanatory Variable	
Long Description	Entire information provided on the dependent variable in text.
Short Description	As found in table.
Data Source	E.g. the Penn World Tables PWT;, World Bank, World Development Indicators WDI; or a data set which some particular researchers came up with.
Lagged	Is the explanatory variable measured in a different (previous) year/month/period than the dependent variable? (Yes=1, No=0)
Logarithmic transformation	E.g. ln, log.; (Yes=1, No=0)
Difference over time	E.g. variable is constructed like: gender gap 2006 "minus" gender gap 2005; (Yes=1, No=0)
Growth rate over time	E.g. variable is constructed like: gender gap 2006 "divided by" gender gap 2005, i.e. a growth rate; (Yes=1, No=0)
Only male component	E.g. male primary enrollment; (Yes=1, No=0)
Only female component	E.g. female primary enrollment; (Yes=1, No=0)
Both male and female component	E.g. male to female primary enrollment ratio; (Yes=1, No=0)
Type of measure	E.g. enrollment, literacy, schooling, years
Construction of variable	How is measure exactly constructed? E.g. logged ratio of men’s and women’s average years of schooling of the population aged 15 and over.
Data	
Description	Full description of data set as found in text.
Data Structure	Cross-section, Cross-country panel, Other (e.g. cross-regional, time-series for single country)

Years	Span of years covered by the particular regression.
Periods	Longitudinal dimensions of the particular regression.
Countries	Cross-sectional dimensions, i.e. numbers of countries covered in the particular regression.
Development level of studied countries	E.g. Middle-income countries, Industrialized countries, Less developed countries, OECD, etc.
Regional focus (if any)	E.g. one or several particular countries; Middle East & North Africa (MENA); Subsaharna Africa (SSA), etc.
Comments	
Method	
Description	Full description of method as found in text.
Method (if studied one time period only)	Cross-section OLS, Cross-section IV, Other cross-section method (specify).
Method (if studied more than one time period)	Pooled cross-section OLS, Random effects, Fixed effects, Panel IV, Other panel method (specify, e.g. GMM, SUR).
Adjustment of error terms	E.g. heteroscedasticity robust standard errors; clustered standard errors etc.