Kyklos

WILEY

KYKLOS

ORIGINAL ARTICLE

The Returns to Education: A Meta-Study

Gregory Clark[1,2,3] | Christian Alexander Abildgaard Nielsen[1]

[1]Department of Economics, University of Southern Denmark, Odense, Denmark | [2]Department of Economic History, London School of Economics and Political Science, London, UK | [3]Center for Economic Policy Research (CEPR), Paris, France

Correspondence: Christian Alexander Abildgaard Nielsen (cniel@sam.sdu.dk)

Received: 9 May 2025 | **Revised:** 26 November 2025 | **Accepted:** 5 January 2026

Keywords: human capital | publication bias | returns to education

ABSTRACT

There have been many studies estimating the causal effect of an additional year of education on earnings. The majority employ administrative changes in the minimum school-leaving age as the mechanism allowing identification. Here, we survey 79 such estimates. However, remarkably, while the majority of these studies find substantial gains from education, a number of well-grounded studies find no effect. The average return from these studies still implies substantial average gains from an extra year of education: an average of 8.2%. But the pattern of reported returns shows clear evidence of publication biases: omission of studies where the return was not statistically significantly above 0, and where the estimated return was negative. Correcting for these omitted studies, the implied average causal returns to an extra year of schooling will be only in the range 0%–3%.

JEL Classification: I26, J24, N3

1 | Introduction

Does more education lead to higher later-life earnings? This is an important question given the strong belief in the importance of human capital among economists, the large public investments in educational programs aimed at increasing levels of education, and current policy debates on further expanding compulsory schooling.[1] Therefore, researchers have sought to determine the causal effects of increasing levels of education, and many conclude that even one extra year of schooling generates substantial earnings gains. Their estimates of substantial causal returns to schooling imply that the extensive provision of public schooling in high-income countries in the modern era should have substantially increased the relative earnings of families with lower social status. Education can potentially be a great equalizer, reducing inequality and increasing social mobility in modern economies.

However, an important question arises from the above. Given the strong established belief in the importance of education, is it easier to publish research supporting this view, while research challenging it faces steeper challenges to publication?

In this paper, we review the literature on returns to education to assess the causal effect of an extra year of compulsory schooling on later-life earnings. Specifically, we conduct a meta-study of 79 estimates of the returns to education derived from increases of the mandated school-leaving age (often described broadly as compulsory schooling laws), or legislation banning child labor under certain ages. Importantly, we also investigate to what extent the reported estimates are affected by publication bias.

In our analysis we focus on two forms of publication bias: (1) That it is easier to publish studies where the return is statistically significantly above 0. Experimental evidence suggests this is a general issue in economics research (Chopra et al. 2024). It has been described as a "null-result penalty", where insignificant results are perceived as less publishable, and as being of lower quality and importance during the publication process. (2) That it is easier to publish studies with positive estimates relative to negative estimates specifically within returns to education literature because many researchers find a negative effect of education inherently implausible.

The first form of publication bias is explored through two tests: first, by determining if there is an association between estimate sizes and standard errors; and second, by investigating if there are signs of *p*-hacking by looking at the distribution of estimates right above and below conventional levels of statistical significance. To evaluate the presence of the second form of publication bias, we investigate whether there is a skewed distribution of estimates. We expected a normal distribution around the average effect (seen in other meta-studies on returns to education), with estimates further away from this average being less likely in both a positive and negative direction. The underlying logic behind each method is elaborated upon in the methods section.

We observe a large average earnings increase of 8.2% for each additional year of compulsory schooling. Using different methods of weighting and correction such as the PET-PEESE procedure suggests a corrected average between 6.0%–6.4%.

However, there is strong evidence in the distribution of reported returns of publication biases in both described forms. Firstly, we uncover a significant and substantially large positive association between the size of estimates and standard errors, suggesting that it is easier to publish significant results. The intercept of this prediction, which is around 3%, suggests that in a world without any publication bias the return to an extra year of schooling could be even lower than the corrected estimates. This association is robust to the inclusion of a number of controls including sample size of the study, gender, country and minimum school-leaving age before the reform, as well as different tests that only utilize subsamples of the data by omitting estimates from older (and thereby potentially less methodologically rigorous) papers, estimates from papers published in less prestigious journals, and finally those estimates that are based on weak instruments potentially leading to a positive bias unrelated to publication bias. Conversely, we find no evidence of *p*-hacking, with a similar number of estimates right above and below conventional significance levels. However, the practice by many authors in the literature of reporting estimates as significant at the 10% level provides a potential reason for lack of *p*-hacking to reach conventional significance levels at 5% and 1%.

We also observe an unexpected nonnormal causal returns distribution that cannot be explained without publication biases. We report a very large number of estimates in the 0%–4% range, 24% of the studies. Given the sampling errors reported in these studies, this peak in the 0%–4% range implies many more studies reporting negative returns than we observe. This suggests an exclusion of papers with negative estimated returns, leading to an artificially high number of positive estimates in the literature. Our interpretation is that underestimations yielding negative returns are more likely to be excluded from publication compared to overestimations yielding higher positive returns.

Taken together, the findings suggest that once you correct published estimates and take omitted studies into account, the implied causal returns to schooling are in the 0%–3% range. We thereby contribute to the literature by revealing strong evidence of publication biases that have led to a general overestimation of the marginal effect of an extra year of education. Furthermore, the constructed dataset of causal estimates can be utilized for further research within the returns to education literature.

A limitation of our study is that we cannot demonstrate specifically how and why this publication bias happens. The researcher himself might discard some studies, choosing not to try to publish findings that show no or a negative effect of an extra year of education. This could be either because he does not believe such findings, interpreting them as a result of errors in the data or method, or simply because he does not believe it possible to publish such a finding. On the other hand, it could be the case that researchers try to publish findings independently of what conclusions they support, but that these papers are rejected in the peer-review process because referees and editors are more likely to interpret the findings as implausible or unreliable. Furthermore, there are many unanswered questions about if and how different organizations such as educational lobbying groups, universities themselves, and government institutions affect the publication process. These are issues which cannot be addressed within the scope of this study.

The paper is organized as follows. Section 2 introduces central methodological challenges in the returns to education literature as well as findings from previous meta-studies focusing on returns to education. Furthermore, we elaborate on how our study contributes to this body of work. In Section 3, we describe the methods utilized to identify publication bias. Section 4 covers how the dataset was constructed. Section 5 contains the main analysis of the average effects of an extra year of schooling, publication bias tests, and a series of robustness checks. Section 6 concludes.

2 | Literature

Estimates of relatively constant returns to schooling across the spectrum of years—primary, secondary, and tertiary—suggest that any extensions to schooling will be socially valuable. Thus an OECD study of the private and social returns to tertiary and “postsecondary nontertiary” education found significant private gains, but also significant public gains that came from increased contribution of taxes to the public treasury (OECD 2013). These estimates also suggest that with appropriate interventions people's social outcomes can be substantially changed. Reflecting this thinking there have been huge investments in education in modern societies. In the United Kingdom in 2017, for example, total investment in education was close to 10% of NNP.

However, the general problem in the estimation of the causal effects of schooling is that people of higher ability choose more schooling, so that the observed earnings gains from schooling likely overestimate the actual returns. A number of strategies have thus been devised to control for ability in estimating the return to schooling, as discussed in Card (1999). Some of those strategies, such as using monozygotic twins to control for ability, despite many high profile publications of such results, have proven to be unsound.[2]

Therefore, Card concluded that the only seemingly reliable methods to estimate the causal effects of education are administrative elements of school systems which cause one cohort of students to get more education than another (Card 1999, 1855). While Card's assessment, where he completely dismisses all other methods, might be too strong, we do agree that methods based on changes to the school system such as compulsory schooling law changes provide the best control for ability bias.[3]

A puzzle has been that such estimates controlling for ability produce estimated returns to education as high or even higher than those from cross-sectional estimates uncontrolled for abilities (Card 1999; Harmon et al. 2003). Since abilities are clearly positively correlated with years of education, this is unexpected. Card reports that the most plausible explanation for this is that "the marginal returns to schooling for certain subgroups of the population – particularly those subgroups whose schooling decisions are most affected by structural innovations in the schooling system – are somewhat higher than the average marginal returns to education in the population as a whole" (Card 1999, 1855).

Despite the clear issue of ability bias in education research, no systematic meta-study has been conducted specifically on causal estimates of economic returns to schooling based on compulsory schooling law changes. Seminal meta-studies of economic returns to schooling only superficially test and discuss publication bias (Ashenfelter et al. 1999; Card 1999), while more recent studies uncritically include uncontrolled and causal estimates together, and some of them focus on one specific country (Montenegro and Patrinos 2014; Churchill and Mishra 2018; Cui and Martins 2021; Ma and Iwasaki 2021). By only including estimates derived from law changes in our meta-study, our analysis is based on a set of more causally reliable estimates that control for ability, relative to previous meta-studies. The similarity in design of the 79 studies we examine provides an ideal terrain for meta-analysis.

There are more comprehensive meta-analyses on related questions of the effects of education. One explores the health effects of education and finds moderate levels of publication bias (Xue et al. 2021). Similarly, a recent working paper investigates issues of publication bias in studies of the effects of free school lunches and reveals that once publication bias is accounted for, the effects are minimal (Ayllón and Lado 2025).

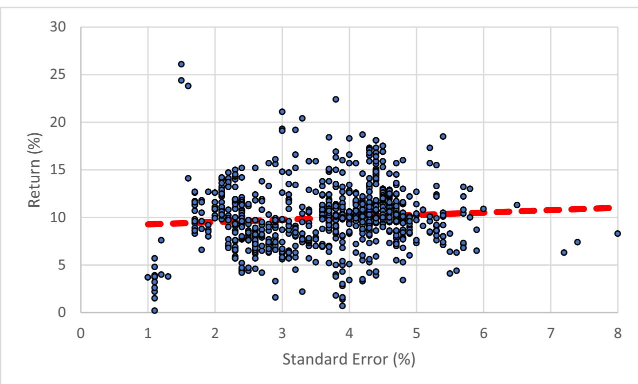


FIGURE 1 | **Gross returns to schooling versus standard error, 1970–2012.** *Source:* Montenegro and Patrinos 2014, Table 1, Columns A and B. [Colour figure can be viewed at wileyonlinelibrary.com]

As argued in the introduction, absent publication biases there should be no connection between the average estimated effect size and the standard error of that estimate. A meta-study of the gross returns to an additional year of schooling (estimated using the Mincer equation which regresses log wage on years of schooling with no controls for ability), for example, reported estimated returns and the standard error of these estimates for 819 estimates (Montenegro and Patrinos 2014). The average return was 10.1%. Figure 1 plots these estimated returns against the standard error of the estimates with the fitted regression line.[4] As can be seen, there is very slight evidence of publication bias in the form of a positive slope of the regression line. However, the slope of this line was not significantly different from 0. Thus, there is no clear evidence of publication bias in this literature.

In the introduction it was also noted that we expected to see a normal distribution around the average estimate size. The above mentioned meta-study of 819 estimates of the gross return to schooling produced just such a normal distribution, shown in Figure 2.[5]

Similarly, Figure 3 shows the distribution of effect sizes, measured in standard deviation units, in a meta-study of 141 randomized control trials of educational interventions, with 271 distinct academic outcomes, where the evaluators were independent of those proposing the intervention, and all the study results were published. These RCTs were commissioned by the UK-based Education Endowment Foundation and the US-based National Center for Educational Evaluation and Regional Assistance, which evaluated interventions aimed at improving academic achievement in kindergarten, primary and secondary education. For a trial to be included in the study, the allocation of children to the intervention and control groups had to be random, and the outcome academic in nature (Lortie-Forgues and Inglis 2019). The mean effect size was surprisingly small—a gain in performance of 0.06 standard deviation units. Here the fit with the normal distribution is not quite as clear, but note

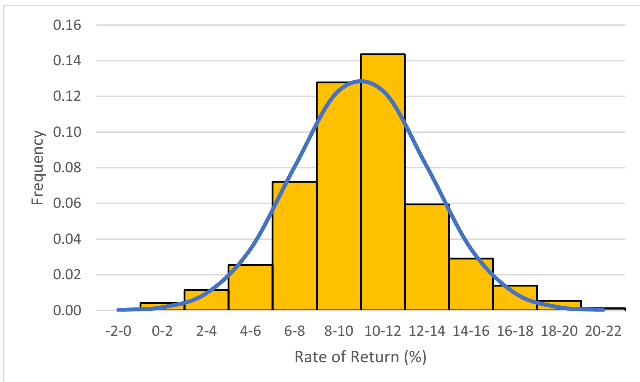


FIGURE 2 | **Distribution of returns to a year of schooling, 1970–2012.** *Note:* Returns estimated using Mincerian regressions of ln (wage) on years of schooling. The normal curve is that based on the mean and standard deviation of reported returns. *Source*: Montenegro and Patrinos 2014, Table 1, Column A. [Colour figure can be viewed at wileyonlinelibrary.com]

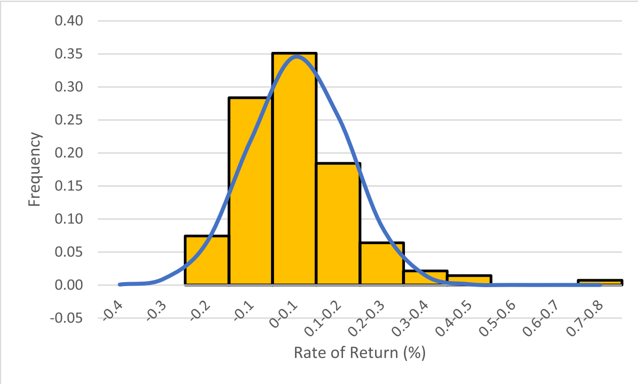


FIGURE 3 | **Distribution of effect sizes from educational RCTs.** *Source:* Lortie-Forgues and Inglis 2019, Figure 1. [Colour figure can be viewed at wileyonlinelibrary.com]

the substantial number of trials which had negative effects on student outcomes.[6]

3 | Methods

Meta-studies have gained increased prominence within economics and are the most objective and statistically rigorous approach to systematic reviews (Stanley and Doucouliagos 2012; Havranek et al. 2020). In meta-studies, researchers combine estimates from multiple studies in order to synthesize results, for example, by determining the average effect of a certain type of policy. Or they seek to discover new patterns, such as the heterogeneous effects of a policy by group or context. Or they seek to detect biases, including publication bias, in the literature.

Publication bias (or publication selection) is the result of editors, reviewers, or researchers having preferences for finding certain results, usually statistically significant results as opposed to null findings (Stanley 2005).[7] Tests for publication bias are especially important because reviewers often dismiss findings based on preconceived notions of theories, potentially leading to biases that cannot be identified or quantified in traditional literature reviews (Stanley 2001).

These biases are potentially very large within returns to education research, where there is a strong established belief in positive returns. Previous research suggests that when almost everyone agrees within a literature is also when the largest distortion biases in what is reported are observed (Stanley and Doucouliagos 2012). Therefore, issues of publication bias are the primary focus of this study.

The most common type of publication bias is favoring significant results relative to null findings. We utilize two tests to assess the presence of this type of publication bias.

First, we investigate the relationship between estimate size and the standard error of the estimate. The rationale behind the test is that "in the absence of publication selection, observed effects should vary randomly around the "true" value independently of the standard error" (Stanley 2005, 321). The implication of this is that a positive association between estimates and errors is suggestive evidence of publication bias. An intuitive way of thinking about this is that it should be just as easy to publish an imprecise estimate close to 0 as an imprecise estimate far from 0. If large estimates are systematically less precise, it suggests insignificant imprecise estimates are being omitted in the publication process while significant estimates are more likely to be published even if very imprecise.

Second, we investigate if there are issues of p-hacking, which can be described as practices that researchers use to generate "better" p-values (Brodeur et al. 2020). We specifically assess whether there is clustering in the distribution of the t-statistics of estimates around 1.96 and 2.58, which corresponds to the 5% and 1% significance levels. Since these conventional levels are arbitrary thresholds used to ease practice and interpretation in research, there is no statistical reason that there should be more estimates on either side of the threshold. If many more estimates are just significant, as opposed to just insignificant, it suggests that researchers engage in p-hacking practices by changing specification or sampling in order to get a significant estimate, which they believe is more likely to get published.

We argue that another type of potential bias in this literature is favoring positive relative to negative estimates independently of significance. Because of the strongly established belief in the importance of human capital for individual outcomes, a negative effect of an extra year of schooling would seem unrealistic to many researchers. Therefore, it is more likely that such results are dismissed by the researcher himself or by referees in the publication process. Furthermore, it is more likely that researchers notice misspecifications or other mistakes that result in a negative estimate, as opposed to mistakes that lead to the overestimation of their already positive estimate.

Therefore, we supplement the above traditional tests with an assessment of the distribution of all estimates. As shown above, in the meta-study of uncontrolled estimates of returns to education (Figure 2), estimates are normally distributed around the average effect. Given that the estimates of the causal returns from schooling are attempting to remove the upwards bias in gross returns caused by on average better students choosing more schooling, we would expect that the causal estimates would show also a normal distribution but with a lower mean. If we instead observe a distribution skewed to one side or the other, it can be interpreted as suggestive evidence that certain estimates are sorted out in the publication process. It is important to note that even if an extra year of education truly never has a negative effect, you would still expect to observe negative estimates due to randomness and errors in sampling and measurement that cause over- or underestimation in the individual estimations.

This approach can be criticized by the fact that we cannot credibly assume that there necessarily is one single "true" effect of an extra year of education. There might be heterogeneous effects based on, among other things, the level of development in the country. However, in the related meta-studies (Figures 2 and 3), estimates do consistently follow a normal distribution centered around the average of estimates, despite heterogeneities also being expected here.

4 | Data

In this review, we follow two simple criteria for inclusion of papers. These criteria were set before the search for papers began. The papers included had to contain:

1. Original data analysis with a causal component based on changes in compulsory schooling laws, or in similar legal or institutional changes, which changed average years of schooling, such as adoption or modification of a child labor law. We also included cases where accidental features such as month of birth, under existing schooling laws, led to differential years of schooling by birth month
2. An adult outcome for the affected children, which can be translated into an estimate of the earnings effect of a change in schooling quantity.

The criteria were chosen to avoid correlational studies, as well as repetitions of the same analysis. Our limited focus on percentage changes to earnings results in a set of highly comparable estimates. The criteria are otherwise unrestrictive in order to include as many relevant papers as possible. Thus, we also included compulsory schooling reforms that were accompanied by changes in the structure of the education system, such as in Sweden in the 1950s where a compulsory schooling extension was accompanied by abolishing tracking based on grades (Meghir and Palme 2005). The assumption here is that the effects of changes in the structure of education were a random shock with respect to changes of time in school. The focus on compulsory education laws or child labor laws means that the results here concern just primary and secondary education.

Both published papers and working papers were included. This could be misleading since we identify publication bias as a major distortion in these estimates. Publication bias can also exist, however, in terms of what papers researchers choose to complete and publicize as working papers. We also show that the results are robust to the exclusion of working papers.[8] We also include a few papers that correct earlier findings through changes to model specification, such as Stephens and Yang 2014 and Fischer et al. 2022. While these replications and corrections are positive in dealing with the publication of spurious findings, we do not believe they solve issues of publication bias in this literature. This is supported by the fact that the issues we observe are as prevalent in recent papers as in older papers. Despite clear improvements in standards of research designs, the issue of publication bias seems to persist.

In the review, we conducted our search for literature in accordance with reporting guidelines for meta-analysis in economics (see Havranek et al. 2020; Irsova et al. 2024). The papers in the review were found primarily by searching on Google Scholar. However, we also utilized Wiley, Elsevir, RePec, and our university's own library system. Our main query was "returns to compulsory schooling," but other search terms were constructed from combinations of the following keywords: "returns to education", "compulsory schooling laws", "compulsory education", and "effects of education." These searches were conducted in September 2024. Many papers were also identified using studies cited in the initial batch of papers we identified, as well as other literature review studies.[9] This extensive process makes us confident that we likely have identified all the relevant papers.

Though most of the papers included below utilize compulsory schooling laws that extend the time children must spend in school, there is a lot of variation in the mechanics of these extensions. Most extensions added an additional year to compulsory school attendance; however, many differed from this pattern, such as Turkey in 1997 that increased years of compulsory schooling from 5 to 8 years (Torun 2018). The typical effect of extensions of compulsory schooling was a sharp discontinuity in educational attainment across one or more years. For example, Figure 4 shows a sharp increase in the fraction of students finishing middle school as a result of the aforementioned Turkish reform.

An important prerequisite for this method to work is that compulsory schooling laws are implemented and significantly affect years of schooling. This is the case for all the schooling laws included in this review. A few papers, however, do have reforms that are relatively weak instruments in some specifications with first stage $F < 10$ (e.g., Buscha and Dickson 2012; Clay et al. 2012; Fang et al. 2012). This is potentially due to a large number of always-takers (children that stay in school past the minimum leaving age irrespective of the reform change) and some never-takers (children taken out of school early by parents despite the reforms).

The returns to education are estimated using different estimation methods, despite a common logic of comparing unaffected and affected cohorts. Most utilize IV estimation, while others employ difference-in-differences or regression discontinuity estimates. No matter the estimation method, these laws affect children that would otherwise have left school as soon as legally possible. They are therefore local average treatment effects (LATE) for children affected by the reforms. This means the studies here are looking at the effects of extended years of education on a grouping of children that on average are of lower academic ability. It is likely that some academically gifted students would leave early for other reasons, such as high opportunity costs if employment is available. However, the observed effects could be quite different from the causal effects of education on earnings for students who choose anyway to stay in school beyond the compulsory leaving age.

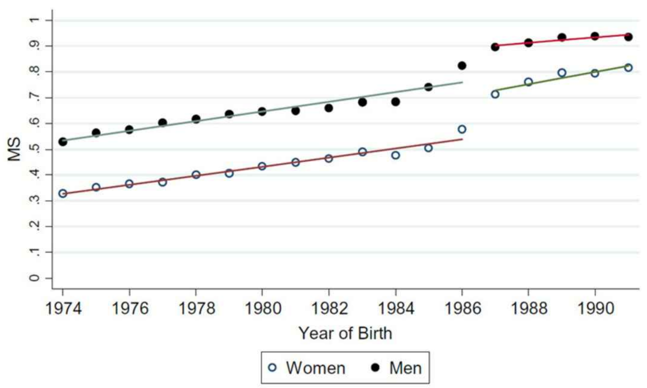


FIGURE 4 | **Educational attainment and compulsory schooling extensions.** *Note*: MS indicates fraction of each birth cohort completing middle school. *Source:* Torun 2018, Figure 2. [Colour figure can be viewed at wileyonlinelibrary.com]

Because papers had multiple estimates, the following criteria were used to select those to include here:

- When multiple estimates were given with no reported "primary model" (e.g., different controls and different bandwidths), we average all causal estimates and their standard errors (excluding robustness checks).
- When a preferred causal estimate was stated by the authors, this is the estimate we report.
- If there are multiple datasets in a paper, we focus on the dataset with the most observations.
- If men and women have separate estimates, each is included separately.
- If the combined sample and only male returns are reported, but not those for women, these reported estimates are taken as two independent estimates.
- When both hourly wage and weekly/monthly/yearly earnings were reported, we report the effect on hourly wages.
- Because all estimates are reported as percentage changes to earnings, no further steps were deemed necessary to standardize or convert estimates to a common metric.
- Similarly, because only independent estimates from each paper are included, no methods are employed to deal with possible dependence across estimates (within study estimates).

We identified 53 papers containing these types of causal estimates of the return to schooling.[10] The papers include estimates from a wide variety of countries. However, there is a clear bias towards developed western democracies, especially the United Kingdom and United States. Specifically, there are 12 papers focusing on the United Kingdom, eight on the United States, three on Sweden, three on Germany, three on Australia, and one or two on a few other western democracies, while two utilize across-country samples within Europe. There are only 14 papers that focus on countries outside the western world, specifically Turkey, China, Indonesia, Taiwan, Thailand, Egypt, Jordan, Argentina, and Mexico.

Most papers (49) focus on reforms that were implemented 1945–2000. Few papers (4) focus on the period before 1945, with the earliest reforms being from 1880.

The total number of reforms in the included papers is 46, most of them 1- or 2-year extensions to minimum school-leaving age.[11] Most reforms are the focus of a specific paper, but some reforms are merely one of many in cross-country samples. Many reforms are used in multiple papers, especially reforms from the United Kingdom as well as the Turkish school reform of 1997 that extended compulsory schooling years from 5 to 8.[12]

5 | Causal Returns to Education

With these procedures the average percentage gain in earnings from an additional year of schooling was 8.2% for 79 independent estimates, from 53 papers.[13] If these estimates are weighed by their precision in a random effects estimation, the gains are reduced to 6.0%, and similarly 6.2% for a fixed effects weighting.[14] Irrespective of just averaging or trying to weight and correct for publication bias, this suggests very large earnings returns to an extra year of compulsory schooling.

Was there any sign of systematic heterogeneity in the returns from education? The reforms that increased required years of schooling, for example, were applied with the current age of required schooling being from 10 to 16 years. Since these reforms mainly affected students of lower academic ability, did the gains from requiring them to get more schooling diminish with the age at which this requirement was imposed?

Figure 5 shows the distribution of estimated returns by minimum school-leaving age before the reform was implemented excluding one outlier estimate of 52% at age 12.[15] Also shown is a fitted regression line of the association between these variables. There is a slight estimated decline in returns with age, as shown in the figure. But that decline is low at −0.93% and not statistically significant at the 5% level ($p = 0.163$).[16] Thus, if there is true underlying heterogeneity in returns to schooling, it has to relate to the details of how the curriculum and its delivery is organized in different school systems.

As mentioned previously, we consider two possible publication biases in this literature. The first is the greater likelihood of an estimate being published when it is statistically significantly different from zero. This will cause biased estimates in two ways. Firstly, results below the standard level of significance will remain unpublished. Secondly, authors will potentially be induced by such publication bias to engage in "*p*-hacking." They will have an incentive to engage in specification search to find the specification and data treatment that creates a result i.e., significant at the conventional 5% or 1% levels. The second bias is the greater likelihood of an estimate being published when the return is greater than zero. Few believe it plausible that the causal effect of education on earnings is negative. Thus, authors may simply abandon studies that produce a negative estimate, or journals will reject such estimates as obviously incorrect.

There is strong suggestive evidence of the first publication bias in the data. Figure 6 shows that the average estimated return rises substantially where the standard error is larger. This suggests a greater likelihood of an estimate being published when it was statistically significantly different from 0. If all estimates were published independent of the estimated return, we would not expect there to be such a gradient.

Figure 7 shows the individual estimates of returns versus standard errors for 65 independent estimates where the standard error was equal to or less than 10%, omitting six outliers.[17][18] Regressing these 65 rates of return on standard errors utilizing OLS-regression with robust standard errors suggests a substantial positive relationship, i.e., significant at the 0.1% significance level ($p = 0.000$).[19] An increase in standard error of 1% is associated with a 1.42% increase of returns to an extra year of education. If all 71 estimates are included, the coefficient is 1.28% ($p = 0.000$). The findings are robust to the inclusion of controls for sample size, GDP per capita (in 2000, current US dollars), minimum school-leaving age before the reform, gender and country, either individually or together (estimates of

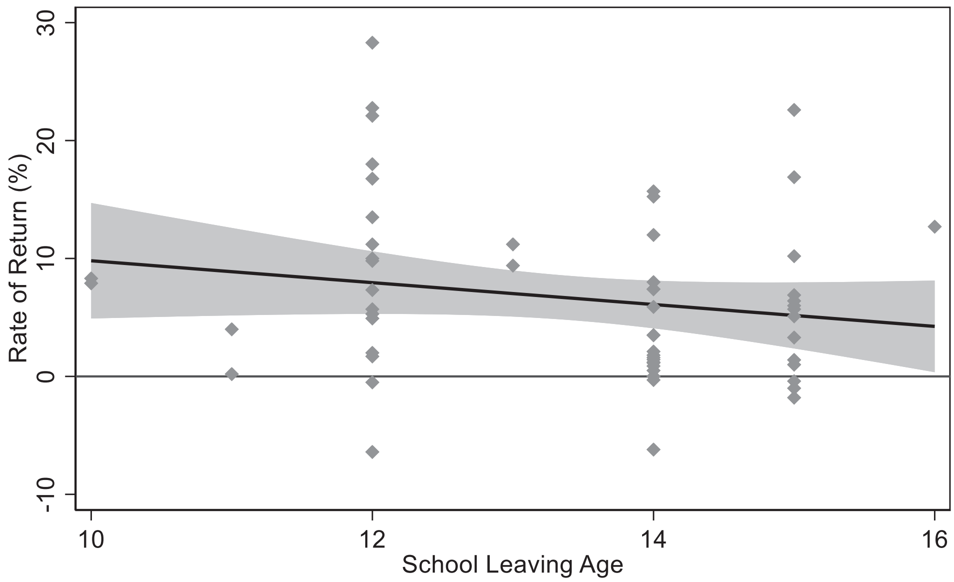


FIGURE 5 | **Rate of return and school-leaving age.** *Note:* $N = 59$. Some observations are excluded because their estimates are based on multiple reforms targeted at different ages. One other observation is omitted from the figure as an outlier, a return of 52%. Significance: Gray bounds indicate 95% confidence intervals. *Source:* Meta-study dataset in online appendix.

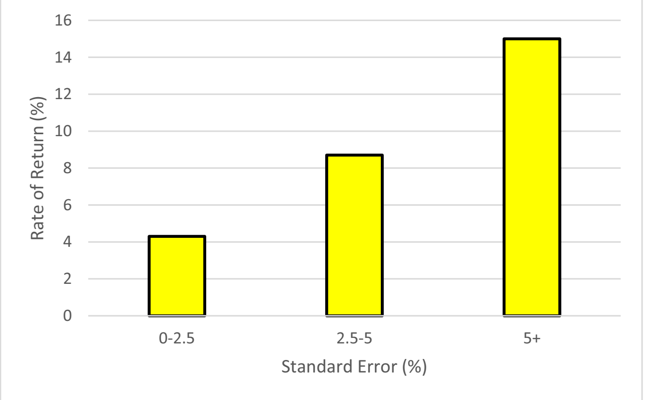


FIGURE 6 | **Rate of return versus standard error.** *Note:* $N = 71$. *Source:* Meta-study dataset in online appendix. [Colour figure can be viewed at wileyonlinelibrary.com]

1.24%–1.36%, $p = 0.000$–0.002).[20] The linear fit to this data suggests that at a standard error of 0, the measured rate of return would be around 3%, as opposed to the 8.2% observed on average. Thus, just this first source of publication bias raises the returns estimated for education by 5%.

The above is only an indicator of publication bias under the assumption of no relationship between the studied effect size and the sample size used to study it (Lau et al. 2006; Simonsohn 2017). With very differing sample sizes (such as in our meta-study), this hypothesis is unlikely to hold. However, in this instance the whole sample consists of similar individuals and outcomes. Furthermore, as mentioned previously, the significance of the association is robust to controlling for sample size (producing an estimate of 1.27%, $p = 0.000$), and the results are robust to restricting the sample to more similar cases (e.g., excluding working papers and papers researching poor-middle–income countries that generally have smaller sample sizes). Lastly, similar issues of differing sample sizes are also true for other meta-studies, where a strong positive association between effect sizes and standard errors is not observed (e.g., the 819 uncontrolled estimates from Montenegro and Patrinos 2014 seen in Figure 1).

Meta-regression analysis offers methods that attempt to account for imprecision and publication bias with the aim of moving the pool of estimates closer to the "true" underlying effect of interest. One widely utilized test is the PET-PEESE procedure (Stanley and Doucouliagos 2014). The Precision Effect Test (PET) regresses estimate sizes on standard errors, while the Precision Effect Estimate with Standard Error (PEESE) uses squared standard errors as the regressor. In both cases, regressions are weighted by the inverse of the sampling variance, meaning more precise studies are given greater weight.

However, the PET-PEESE procedure has recently been criticized as producing uncertain corrections and inflated standard errors, leading to false negatives, especially when the number of studies is small (Pustejovsky 2017; Doucouliagos et al. 2018; Bartoš et al. 2022). Using preregistered multiple-laboratory replications to provide a context without publication bias, Kvarven et al. 2020 conclude that meta-analytic estimates from PET-PEESE have corrected estimates that are inflated by 31% relative to the replications, and the greater imprecision leads to a false-negative rate of 50%.

When we utilize this method, the results show weak signs of publication bias that are very model dependent. Using the same data with outliers removed, the intercept of the regression suggests a true effect of 6.1%–6.4% to an extra year of education and an insignificant effect of publication selection, specifically $p = 0.498$ for PET and $p = 0.288$ for PEESE. This can be interpreted as suggestive evidence that the pattern we observe above is not the result of publication bias but of actual heterogeneity in the true effect. Once we account for many potential

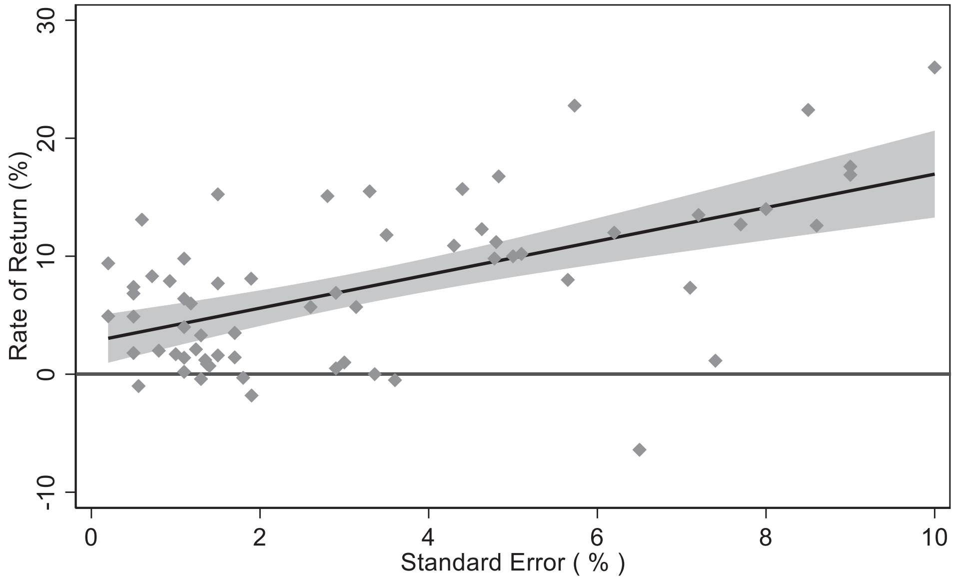


FIGURE 7 | **Rate of return versus standard error.** *Note:* $N = 65$. Six outliers with standard errors above 10% omitted. Significance: Gray bounds indicate 95% confidence intervals (robust standard errors). *Source:* Meta-study dataset in online appendix.

heterogeneities by including previously mentioned controls, the most conservative PET test is insignificant ($p = 0.103$), while the PEESE test becomes significant ($p = 0.026$). Small changes to specification greatly affect these results. If outliers are included in the controlled regression, both the PET and PEESE test show significant support for publication bias (respectively, $p = 0.031$ and $p = 0.007$). This is the only evidence that challenges our interpretation of the presence of publication bias. But as noted, the PET-PEESE procedure has now been challenged as a valid test.

There was a much larger average effect of 11.3% in poor- to middle-income countries compared to 6.6% for high-income countries.[21] This suggests that there could be larger returns to education in less developed countries with a low base level of education. However, the average standard error was also much bigger for the poorer countries (6.6%) compared to high-income countries (3.3%). Potentially then, the higher observed returns in poor- to middle-income countries could also be the result of publication biases in favor of statistically significant results. This argument is supported by the fact that both GDP per capita and a poor-middle–income country dummy have an insignificant association with estimated effect size when controlling for the standard errors of the estimate (respectively, a coefficient of 0.00002, $p = 0.756$, and -0.12, $p = 0.943$).

As Card (1999) noted, IV estimates tend to be as large or larger than the OLS estimates, despite the OLS estimates being upward biased by the selection into education of those of higher ability. More recent studies conclude that IV estimates are as much as 36% higher than OLS estimates on average (Soon and Lim 2024; Patrinos and Psacharopoulos 2025). We observe a similar pattern. 49 IV estimates had a corresponding OLS estimate. In 31 of the 49 cases (63%), the IV estimate was bigger. These IV estimates thus average 9.3% compared to 6.9% for the OLS estimates for the same population, which is 35% higher. Card explained this through a higher return to education among those affected by extensions of compulsory schooling than among those choosing additional schooling. But this implies that the estimated return to education using compulsory schooling laws will reveal very little about the return to education for most educational investments.

A more likely explanation for why the IV estimates here are larger than the OLS estimates would again be publication bias. The IV estimates will have larger standard errors than OLS estimates. If the papers they appear in are more likely to be published where the IV estimate is statistically positive, then there will be a selection towards higher IV estimates that does not apply to the OLS estimates, helping explain the higher average IV returns.

Recent research provides another potential explanation for the larger IV estimates, which challenges our explanation based on publication bias. Keane and Neal (2023, 1626) argue that if the OLS bias is positive (likely the case due to selection into schooling based on ability) then "the 2SLS t-test will have inflated power to find false positive effects and little power to detect true negative effects". Because of this, the strong association between estimate size and standard errors might simply be due to the prevalence of many weak instruments in the literature, and not selection for publication based on significant findings.

We empirically address this issue by excluding instruments with a weak first stage from our analysis. Many researchers use $F > 10$ as the rule of thumb for a strong instrument. Keane and Neal (2023) argue that the issue of instruments being more likely to find false positive effects is also very likely for instruments in the range of $F = 10$–20. Lee et al. (2022) describe that this issue is only truly unlikely with F-statistics as high as 104.7. Therefore, we replicate our main regression that regresses standard errors on estimate sizes, excluding outliers with standard errors above 10% (visualized in Figure 7), including only IV estimates with first stage F above 20 and 104.7.[22]

Excluding instrumental estimates with F-statistics below 20 reduces our sample from 65 to 57. The coefficient increases from

1.42% to 1.63% ($p=0.000$).[23] This suggests that our findings cannot be satisfactorily explained by a positive bias due to weak instruments. However, if we use the very restrictive level of 104.7, reducing our sample further to 39 estimates, the association is weaker at 0.70% and only marginally significant ($p=0.049$).[24] This lends support to the argument by Keane and Neal (2023) because it suggests a stronger positive bias among weak instruments in our sample. Therefore, we support the authors' calls for a more critical approach to IV analysis, where applied researchers only use instruments with F-statistics far above the conventional level of 10. However, even with this very restrictive sample that excludes weak instruments, we still observe a significant association, suggesting that it is easier to publish estimates showing significant returns to an extra year of compulsory schooling.

There is, however, no sign in response to this first bias of further bias through p-hacking. To test for this, we looked at a bandwidth of 0.2 around a t-statistic of 1.96 and 2.58 (5% and 1% confidence intervals). What were the relative numbers of t-statistics in this band above and below 1.96 and 2.58? The answer is 6 above versus 6 below for 1.96 and 4 versus 3 for 2.58. Thus, there is no sign of p-hacking in the results.

This is not in line with recent literature that finds high levels of p-hacking in economics (Brodeur et al. 2020). However, it is difficult to conclude much from the test due to the low number of observations close to the thresholds. Furthermore, many papers in the literature interpret an estimate significant at the 10% level as a significant effect of an extra year of schooling. This reduces the incentives to engage in p-hacking to reach the conventional 5% and 1% levels.

As noted above in the methods section, based on the cases of reported returns to education (Figures 2–3), we would expect, absent publication biases, that the estimated causal returns to education would follow a normal distribution. The average return in noncausal estimates reported in Montenegro and Patrinos 2014 of 10.1% is, after all, not much above the average of 8.2% for the causal estimates in our analysis. This is assuming the two sources of variation in reported returns to education are a normally distributed underlying rate of return, combined with sampling error. Simulation using the reported standard errors of estimates suggests sampling error will explain half the reported variance in returns. However, as Figure 8 shows, estimated returns to education have a highly asymmetric distribution with a strong peak in measured returns in the 0%–4% range, despite the mean measured return being 8.2%. The chance of getting the observed 19 (24% of estimates) or more returns in the 0%–4% range, compared to the expected 10.5 with the given normal distribution, is only 0.5%.

One possible interpretation of Figure 8 is that because of a strong belief that the true returns to education cannot be negative, there is omission from published studies of those that find negative returns. As can be seen in Figure 9, the reported studies with a positive return to education in fact approximate well to the upper half of a normal distribution of returns to education centering close to 0, and with the same variance as these positive return studies.

A challenge to this interpretation is that the clustering of estimates around 0 is somehow caused by weak instruments leading to a downward bias. However, the above analysis shows that

FIGURE 8 | **Estimated rates of return versus a normal distribution**. *Note:* $N=79$. Estimated rates of return to a year of education in the meta-study in range −12%–28%, versus normal distribution with same mean and variance. *Source:* "Data for figs. 8–10.xlsx" in online appendix. [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 9 | **Estimated rates of return versus a normal distribution with mean 0.** *Note:* $N=79$. Estimated rates of return to a year of education in the meta-study in range -12%–32%, versus normal distribution with mean 0, and same variance for positive returns. *Source:* "Data for figs. 8–10.xlsx" in online appendix. [Colour figure can be viewed at wileyonlinelibrary.com]

if weak instruments bias the estimates in any direction, it is upward. Furthermore, most of the large estimates were IV estimates, with a mean of 10.1% (57 estimates) relative to a mean of 3.4% for difference-in-differences and regression discontinuity estimates (22 estimates).

It has also been suggested that the heavy concentration of returns in the range just above 0% could be the product of the underlying returns having a log normal distribution, which would imply no true returns falling in the range below 0.[25] You would expect, however, that since the average causal return estimate (8.2%) is close to the average gross return estimate (10.1%), then gross returns would also display a log normal distribution. As seen in Figure 2, those estimates show a normal distribution.

However, when we plot an underlying log normal distribution with the same sampling error as is reported in the studies (orange bars in Figure 10), again this distribution does not fit the

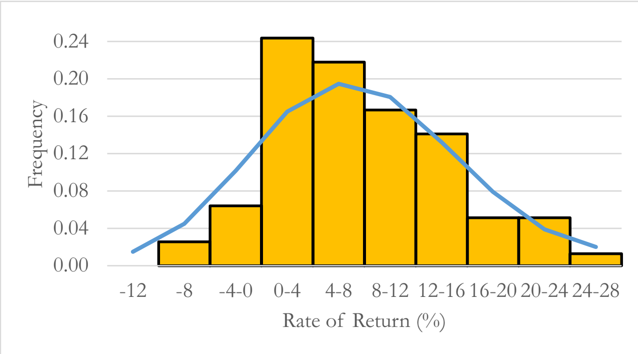


FIGURE 10 | **Estimated rates of return versus a log normal distribution.** *Note:* $N = 79$. Estimated rates of return to a year of education in the meta-study in range -12%–28%, versus log normal distribution with same mean and variance. *Source:* "Data for figs. 8–10.xlsx" in online appendix. [Colour figure can be viewed at wileyonlinelibrary.com]

reported returns. Because of the importance of sampling error in generating the observed distribution, the resulting distribution should deviate little from a normal distribution. The peak of the returns generated by the log normal distribution is 4%–8%, not the 0%–4% observed in the studies. The log normal distribution also still overpredicts the number of observed negative returns.

Thus in addition to clear evidence of omission of positive estimates that are statistically not different from 0, there is also sign of omission of estimates of negative returns that would be expected given sampling errors, and the cluster of estimates that fall in the 0%–4% range. In addition to just simple omission of negative estimates, this could suggest a different form of hacking, which is best described as 0-hacking. Researchers might be induced to alter specifications to get estimates just above 0. This implies that the average causal rate of return may be even lower than the 3% suggested above where we just consider the first set of omissions.

An issue with arguing that the conducted meta-study is evidence of current issues of publication bias is the inclusion of older papers. Could the results simply be driven by old research that was less methodologically rigorous? This is clearly a possibility with the growing importance of the "credibility revolution" and constantly evolving identification and estimation methods, alongside the increasing size and quality of datasets. However, while these developments have likely contributed to a reduction of many forms of bias, the incentives leading to publication bias remain unchanged. Researchers and publishers having a preference for or being more likely to believe significant and positive findings is not unique to old research, and no number of new difference-in-difference estimators will solve this underlying issue. In a worst case scenario for publication bias, these developments might simply provide further researchers degrees of freedom to reach results that align with a priori expectations.[26] Furthermore, evidence does not support the conclusion that publication bias is only present in old papers. Average returns are similar for new and old papers, for example, papers published in 2015–2024 averaged a 7.5% return (43 estimates), while papers in 1991–2014 averaged only modestly higher at 9.1% (36 estimates).[27] Furthermore, standard errors are only marginally larger in older papers with an average of 4.5% (33 estimates) relative to 4.3% (38 estimates) in newer papers. Also, the presence of negative estimates is similar with three in 2015–2024 and five in 1991–2014. However, the most convincing evidence is that among new papers we still observe a significant positive relationship between the size of estimates and standard errors, and the distribution has a similar pattern with many observations just above 0.[28] We also replicate the controlled regression between estimate sizes and standard errors, adding publication year as an extra control, which changes the estimate from 1.24% to 1.25%. Lastly, the color coded funnel plot (Figure F1 in Appendix F) shows that estimates from old and new papers are close to equally represented outside the pyramid (indicating publication selection).[29] This is unfortunately not a story of a lack of methodological rigor in the past, but of ongoing publication bias in the present.[30]

We also investigated whether there is a relationship between the size of estimates and number of citations. Do large and significant positive returns to education in a paper elicit more citations of that paper? This supplementary test can reveal if large positive returns also garner more attention after publication. Once we control for year of publication, a dummy indicator of the paper that is considered the pioneering study (Angrist and Krueger 1991), and use the log of citations to reduce the influence of outliers, we find no relationship between the estimated effect size and citation frequency. Specifically, an estimate of -0.01 ($p = 0.578$), meaning a very modest and insignificant decrease in citations per each 1% increase in returns to a year of compulsory schooling. We also tested if there was a relationship between citation frequency and the size of standard errors. Our expectation was that papers producing more precise estimates should be judged more favorably and garner more citations. Similarly, there was no relationship here with an estimate of -0.05 ($p = 0.071$). Neither bigger estimated returns, nor more precise estimates, garnered more citations.

Thus, whatever is driving the observed publication biases for estimates of the return to schooling, it is not that estimates with smaller effect sizes once published are regarded as dubious by the research community.

Lastly, it is important to note that despite our primary purpose of compiling the data was to map out the most causally reliable estimates of returns to education as well as explore issues of publication bias, the data opens up for other avenues of education research. For example, when researching heterogeneous effects, the inclusion of estimates from multiple papers (samples) might reduce the likelihood of randomness in a specific sample explaining differences in outcomes between groups.

One interesting question is if returns are different for men and women. Returns might differ by gender for several reasons. Potentially, returns are lower for women if increased education is not sufficiently supplemented with access to high-paying occupations for women once they enter the labor market. Returns might conversely be higher for women if there are more high-paying alternatives for men that do not require much education.

If we simply observe the average return for all independent estimates by gender, there is some suggestive evidence of gender differences. However, returns are high for both men and women. The average return for women is 8.01% ($N = 23$) and 10.01% ($N = 32$) for men.[31] However, the difference is not significant when a gender dummy is regressed on returns ($p = 0.450$). When we include controls for the country's GDP, sample size of the study and the minimum school-leaving age before the reform, the lower returns for women become more pronounced with an estimate of −5.27% but are still insignificant at the 5% level ($p = 0.148$). To conclude on this brief illustration, there is weak evidence of gender differences. However, nothing can be concluded with statistical certainty due to the small sample size.

We encourage further research on the heterogenous effects of education using metadata, either by expanding upon our data or with the creation of new datasets. One issue with researching heterogeneity with our dataset, as is clear in the above illustration, is the relatively small sample size. This presents a trade-off to researchers when collecting their metadata, where the right decision is based on the specific research question. Including many estimates will make heterogeneity tests more plausible but will often come at the expense of having a more causally reliable and comparable set of estimates.

6 | Conclusion

In the previously cited meta-study of returns to education by Ashenfelter et al. (1999), they also report evidence of publication bias with IV estimates of the returns to a year of schooling. They thus report a significant positive correlation between the estimated effect size and the standard error. Despite the strong correlation between the return estimate and the standard error, their method of correction only reduces the average IV return estimate from 8.6% to 8.1%, an inconsequential amount (Ashenfelter et al. 1999).

In this meta-study of 79 estimates of the causal effects of education on earnings, we report a similar effect to that earlier study: an 8.2% gain in earnings for each year of education.[32] We also find clear evidence of significant publication biases. Studies passing the filter of publication were more likely to have coefficients statistically significantly greater than 0. Furthermore, we also observe strong evidence that studies with returns estimated at less than 0 were more difficult to publish. Correcting the first publication bias by looking at the average return on the studies with the lowest standard errors would suggest an average causal return to education of no more than 3%, compared to the reported 8.2%. Independent of this, correcting for the omission of returns less than 0 implies an overall causal return to education only in the 0%–3% range. Our findings are in line with experimental evidence showing a "null result penalty", where insignificant results in economic research are perceived as less publishable, and as being of lower quality and importance (Chopra et al. 2024), and with Xue et al. 2021 that find moderate levels of publication bias when reviewing effects of education on health. Another meta-study similarly found that estimates in economics literature in general are overestimated in close to 80% of cases (Ioannidis et al. 2017).

There are a large number of other papers on extensions of compulsory education that measure other outcomes such as employment rates, higher education attainment, mortality rates, fertility, and criminality. However, for none of these outcomes were there sufficient numbers of papers with a given comparable outcome such that we could carry out the same analysis as above. However, given the publication biases we observe with earnings, we suspect that papers showing increased compulsory education leading to higher employment, more university education, lower mortality, and decreased criminality are all likely to have faced similar problems with publication bias.

Author Contributions

Gregory Clark: conceptualization, methodology, formal analysis, writing (original + revisions). **Christian Alexander Abildgaard Nielsen:** conceptualization, data curation, methodology, formal analysis, writing (original + revisions).

Funding

Clark's professorship and Nielsen's PhD-position at the University of Southern Denmark are partly funded by Clark's Danish National Research Foundation (DNRF) chair grant. DNRF had no direct involvement in this specific project: https://dg.dk/en/centers/dnrf-chair-professor-gregory-clark/.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Underlying data and online appendix is openly available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BUOUNW (Clark and Nielsen 2025).

Endnotes

[1] For example, Sweden's government is considering requiring an extra year of compulsory schooling (Bryant 2024).

[2] In Appendix A, we discuss why twin estimates are unreliable.

[3] An increasing number of studies investigate returns to higher education based on other causal mechanisms, for example, by utilizing distance to college (Card 1995), changes in criteria for higher education (Brunello and Miniaci 1999), and grade thresholds (Zimmerman 2014). However, there is still some level of selection on ability in terms of who chooses to pursue higher education, when it becomes more accessible. This issue of opting in our not is not present when utilizing compulsory schooling laws.

[4] The regression line was fitted including an indicator for returns in high-income countries.

[5] However, there are signs even in this case of some publication bias. There were no negative estimates of returns. If there was symmetry with the upper tail of the estimates, there should have been six such estimates. Correcting for this likely omission the average return was 10% instead of 10.1%.

[6] Given that these were RCTs, independently evaluated, we would expect that the effect size found would be, as in Figure 1, largely independent of the standard error. However, presumably just by chance, there was a significant positive association between the effect size and the standard error.

[7] Stanley and Doucouliagos 2014 argue that "reporting bias" is a more accurate term, since the selection is also performed by researchers

that are aware of preferences for statistically significant findings, before the publication process.

[8] Results excluding working papers and papers from "less prestigious journals" (proxied by journal impact factor in 2022) can be replicated using the do-file in the online appendix. Excluding working papers reduces the average return from 8.2% (79 estimates) to 7.2% (68 estimates).

[9] Following the recommendations of Page et al. 2021, we document our process of searching for and sorting out papers in greater detail in the PRISMA diagram in Appendix B.

[10] A list of all papers is reported in Appendix C.

[11] Some papers utilize variations in multiple law changes across different states in the United States (e.g., Angrist and Krueger 1991; Acemoglu and Angrist 2000; Li 2023), Canada (Oreopoulos 2006b), and Australia (Powdthavee 2013). Each of these batches of laws are counted as one reform in the above summary.

[12] A very brief description of all reforms is available in Appendix D.

[13] The average return was 8.5% from 92 estimates, where not all were independent, again from 53 papers. Note, however, that of these 92 estimates in total, only 49 (53%) were statistically significant at the 5% level. Definitions, means, and standard deviations for variables used in the analysis as well as a codebook for other variables coded is available in Appendix E.

[14] The weights in the random effects model vary from 0.08% to 2.19% meaning no single or a few estimates explain a majority of the mean. Conversely, in the fixed effects model, two observations are given a weight of 27.52% (55.04% in total between the two).

[15] Some reforms increased required years of required school attendance. In these cases, we estimated the age of the students affected assuming formal education began at age 6.

[16] Given the above argument, we assume any heterogeneity would follow a linear pattern of a continued decrease. If we make no such assumptions and treat school-leaving age as a categorical variable, we do observe significant differences for four of the six categories relative to the reference group (age 10). However three of these categories only have 1–2 observations, and it follows no discernible logical pattern, for example, with age 14 having a stronger negative estimate than age 15. This suggests that there is no strong evidence for effect heterogeneity by school-leaving age.

[17] Eight other estimates are not included in Figures 6–7, because some papers did not report standard errors.

[18] This is typically visualized as a funnel plot. A funnel plot of the same data is illustrated in Figure F1 in Appendix F, which is based on Egger et al. 1997 that weights estimates by their precision. Our conclusions remain unchanged, when observing the data in this way. This is also the case when we include the controls described above (both individually and together) to account for possible effect heterogeneity. The color coding of observations are based on estimates being from old or new papers, which is used in a later discussion.

[19] Since we are just testing whether the slope of the line is positive, we use a one-tailed test here and below. However, the estimate is also significant at 0.1% with a two-tailed test.

[20] Note, however, that this also greatly decreases the sample sizes (from 71 to between 47 and 57).

[21] Poor to middle income: countries with GDP per capita below current 10.000 US dollars in the year 2000. For example, China, Jordan, Argentina, Turkey, and Poland.

[22] Note that some studies do not report F-statistic meaning they are omitted from this specific estimation. Many authors report the coefficient and standard error of the first stage but not the F-statistic. In these cases, an approximate F-statistic is calculated using $F = \frac{\hat{\beta}}{SE\left(\hat{\beta}\right)}$.

[23] This association is robust to the inclusion of controls ($\beta = 1.37\%, P = 0.000$).

[24] This finding is less robust. If we include the very unprecise outliers with standard errors above 10%, the relationship is insignificant ($\beta = 0.58\%, P = 0.208$). Similarly, the association becomes insignificant with the inclusion of controls resulting in a sample size of only 24 ($\beta = 1.00\%, P = 0.242$). A less robust finding is to be expected here, since the sample now consists primarily of very precise estimates, where publication bias is not expected (Stanley and Doucouliagos 2014, 63).

[25] Rachael Meager made this criticism in a posting on X.

[26] There are, for example, more opportunities to pick among a larger set of estimation methods (as more new techniques are developed), subsamples and controls (due to growing size of datasets and number of variables in the datasets), weighting methods, as well as choice of bandwidths, and more.

[27] In the do-file, we also provide code for constructing a figure showing all estimates by publication year, which signifies no meaningful differences in the correlation across time.

[28] See the do-file in online appendix for replication.

[29] The choice of new papers as starting in 2015 can fairly be criticized as arbitrary. The threshold was simply chosen to have sufficient sample sizes for both periods. See do-file for replication of the main regression with later years as the starting point for "new" papers, leading to weaker significance (2017, $p = 0.073$; 2019, $p = 0.086$; 2021, and $p = 0.000$). However, the very small sample sizes make these tests uncertain.

[30] As previously mentioned in Footnote 8, we also conducted the same exercises excluding working papers and papers from "less prestigious journals" (proxied by journal impact factor in 2022) from the sample, which similarly did not change our conclusions. See do-file in online appendix for replication.

[31] The number of estimates only add up to 55 because some papers exclusively report a "full" sample, making gender decomposition impossible.

[32] If we exclusively include IV estimates it would be a 10.1% gain from 57 estimates.

References

Aakvik, A., K. G. Salvanes, and K. Vaage. 2010. "Measuring Heterogeneity in the Returns to Education Using an Education Reform." *European Economic Review* 54: 483–500.

Acemoglu, D., and J. Angrist. 2000. "How Large Are Human-Capital Externalities? Evidence From Compulsory Schooling Laws." *NBER Macroeconomics Annual* 15: 9–59.

Albarrán, P., M. Hidalgo-Hidalgo, and I. Iturbe-Ormaetxe. 2020. "Education and Adult Health: Is There a Causal Effect." *Social Science and Medicine* 249, no. 112830: 1–9.

Alzúa, M. L., L. Gasparini, and F. Haimovich. 2015. "Education Reform and Labor Market Outcomes: The Case of Argentina's Ley Federal De Educación." *Journal of Applied Economics* 18, no. 1: 21–43.

Amin, V. 2011. "Returns to Education: Evidence From UK Twins: Comment." *American Economic Review* 101, no. 4: 1629–1635.

Angrist, J. D., and A. B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106, no. 4: 979–1014.

Ashenfelter, O., C. Harmon, and H. Oosterbeek. 1999. "A Review of Estimates of the Schooling Earnings Relationship, With Tests for Publication bias." *Labour Economics* 6, no. 4: 453–470.

Ashenfelter, O., and A. B. Krueger. 1994. "Estimates of the Economic Return to Schooling From a New Sample of Twins." *American Economic Review* 84, no. 5: 1157–1173.

Ashenfelter, O., and C. Rouse. 1998. "Income, Schooling and Ability: Evidence From a New Sample of Identical Twins." *Quarterly Journal of Economics* 113, no. 1: 253–284.

Aydemir, A., and M. G. Kirdar. 2017. "Low Wage Returns to Schooling in a Developing Country: Evidence From a Major Policy Reform in Turkey." *Oxford Bulletin of Economics and Statistics* 79, no. 6: 1046–1086.

Ayllón, S., and S. Lado. 2025. "The Causal Impact of School-Meal Programmes on Children in Developed Economies: A Meta-Analysis." IZA Discussion Paper No. 18042.

Bartoš, F., M. Maier, D. S. Quintana, and E.-J. Wagenmakers. 2022. "Adjusting for Publication Bias in JASP and R: Selection Models, PET-PEESE, and Robust Bayesian Meta-Analysis." *Advances in Methods and Practices in Psychological Science* 5, no. 3: 1–19. https://doi.org/10.1177/25152459221109259.

Behrman, J. R., and M. R. Rosenzweig. 1999. "Ability Biases in Schooling Returns and Twins: A Test and New Estimates." *Economics of Education Review* 18, no. 2: 159–167.

Behrman, J. R., M. Rosenzweig, and P. Taubman. 1994. "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment." *Journal of Political Economy* 102, no. 6: 1131–1174.

Bhuller, M., M. Mogstad, and K. G. Salvanes. 2017. "Life-Cycle Earnings, Education Premiums, and Internal Rates of Return." *Journal of Labor Economics* 35, no. 4: 993–1030.

Bonjour, D., L. F. Cherkas, J. E. Haskel, D. D. Hawkes, and T. D. Spector. 2003. "Returns to Education: Evidence From UK Twins." *American Economic Review* 93, no. 5: 1799–1812.

Brodeur, A., N. Cook, and A. Heyes. 2020. "Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review* 110, no. 11: 3634–3660.

Brunello, G., M. Fort, and G. Weber. 2009. "Changes in Compulsory Schooling, Education and the Distribution of Wages in Europe." *Economic Journal* 119, no. 536: 516–539.

Brunello, G., and R. Miniaci. 1999. "The Economic Returns to Schooling for Italian Men. An Evaluation Based on Instrumental Variables." *Labour Economics* 6: 509–519.

Bryant, M.. 2024. "Swedish Children to Start School a Year Earlier in Move Away From Play." The Guardian. Retrieved December 2025. https://web.archive.org/web/20250725062539/https://www.theguardian.com/world/2024/sep/19/swedish-children-to-start-school-a-year-earlier-six.

Buscha, F., and M. Dickson. 2012. "The Raising of the School Leaving Age: Returns in Later Life." *Economic Letters* 117: 389–393.

Buscha, F., and M. Dickson. 2015. "The Wage Returns to Education Over the Life-Cycle: Heterogeneity and the Role of Experience." IZA Discussion Paper No. 9596.

Card, D. 1995. "Using Geographic Variation in College Proximity to Estimate the Returns to Schooling." NBER Working Paper 4483.

Card, D. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, edited by O. Ashenfelter and D. Card, vol. 3A. Elsevier.

Chen, Y., S. Jiang, and L.-A. Zhou. 2020. "Estimating Returns to Education in Urban China: Evidence From a Natural Experiment in Schooling Reform." *Journal of Comparative Economics* 48: 218–233.

Chib, S., and L. Jacobi. 2016. "Bayesian Fuzzy Regression Discontinuity Analysis and Returns to Compulsory Schooling." *Journal of Applied Econometrics* 31: 1026–1047.

Chopra, F., I. Haaland, C. Roth, and A. Stegmann. 2024. "The Null Result Penalty." *Economic Journal* 134, no. 657: 193–219.

Churchill, S. A., and V. Mishra. 2018. "Returns to Education in China: A Meta-Analysis." *Applied Economics* 50, no. 54: 5903–5919.

Clark, D. 2023. "School Quality and the Return to Schooling in Britain: New Evidence From a Large-Scale Compulsory Schooling Reform." *Journal of Public Economics* 223: 1–10.

Clark, G., and C. A. A. Nielsen. 2025. "Clark & Nielsen—The Returns to Education: A Meta-Study V5." Harvard Dataverse. Retrieved December 2025. https://dataverse.harvard.edu/dataset.xhtml%3FpersistentId%3Ddoi:10.7910/DVN/BUOUNW.

Clay, K., J. Lingwall, and M. S. Jr. 2021. "Laws, Educational Outcomes, and Returns to Schooling Evidence From the First Wave of U.S. State Compulsory Attendance Laws." *Labour Economics* 68: 101935.

Clay, K., J. Lingwalland M. Stephens Jr.. 2012. "Do Schooling Laws Matter? Evidence From the Introduction of Compulsory Attendance Laws in the United States." NBER Working Paper 18477.

Cui, Y., and P. S. Martins. 2021. "What Drives Social Returns to Education? A Meta-Analysis." *World Development* 148: 105651.

Delaney, J. M., and P. J. Devereux. 2019. "More Education, Less Volatility? The Effect of Education on Earnings Volatility Over the Life Cycle." *Journal of Labor Economics* 37, no. 1: 101–137.

Devereux, P. J., and R. A. Hart. 2010. "Forced to Be Rich. Returns to Compulsory Schooling in Britain." *Economic Journal* 120: 1345–1364.

Dickson, M. 2013. "The Causal Effect of Education on Wages Revisited." *Oxford Bulletin of Economics and Statistics* 75, no. 4: 477–498.

Dolton, P., and M. Sandi. 2017. "Returning to Returns: Revisiting the British Education Evidence." *Labour Economics* 48: 87–104.

Domnisoru, C. 2021. "Heterogeneity Across Families in the Impact of Compulsory Schooling Laws." *Economica* 88: 399–429.

Doucouliagos, H., M. Paldam, and T. D. Stanley. 2018. "Skating on Thin Evidence: Implications for Public Policy." *European Journal of Political Economy* 54: 16–25.

Eble, A., and H. Feng. 2019. "Does Primary School Duration Matter? Evaluating the Con-Sequences of a Large Chinese Policy Experiment." *Economics of Education Review* 70: 61–74.

Egger, M., G. D. Smith, M. Schneider, and C. Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *BMJ* 315, no. 7109: 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Elsayed, A. Z., and O. Marie. 2020. "Less School (Costs), More (Female) Education? Lessons From Egypt Reducing Years of Compulsory Schooling." Tinbergen Institute Discussion Papers.

Fang, H., K. Eggleston, J. A. Rizzo, S. Rozelle, and R. J. Zeckhauser. 2012. "The Returns to Education in China: Evidence From the 1986 Compulsory Education Law." *NBER Working Paper* 18189.

Fischer, M., G. Heckley, M. Karlsson, and T. Nilsson. 2022. "Revisiting Sweden's Comprehensive School Reform: Effects on Education and Earnings." *Journal of Applied Econometrics* 37: 811–819.

Fischer, M., M. Karlsson, T. Nilsson, and N. Schwarz. 2020. "The Long-Term Effects of Long Terms—Compulsory Schooling Reforms in Sweden." *Journal of the European Economic Association* 18, no. 6: 2776–2823.

Gerritsen, S. 2014. "Zero Returns to Compulsory Schooling. Is It Certification or Skills That Matter?" CPB (Netherlands Bureau for Economic Policy Analysis) Discussion Paper 293.

Grenet, J. 2013. "Is Extending Compulsory Schooling Alone Enough to Raise Earnings? Evidence From French and British Compulsory

Schooling Laws." *Scandinavian Journal of Economics* 115, no. 1: 176–210.

Harmon, C., H. Oosterbeek, and I. Walker. 2003. "The Returns to Education: Microeconomics." *Journal of Economic Surveys* 17, no. 2: 115–156.

Harmon, C., and I. Walker. 1995. "Estimates of the Economic Return to Schooling for the United Kingdom." *American Economic Review* 85, no. 5: 1278–1286.

Havranek, T., T. D. Stanley, H. Doucouliagos, et al. 2020. "Reporting Guidelines for Meta-Analysis in Economics." *Journal of Economic Surveys* 34, no. 3: 469–475.

Hicks, D., and H. Duan. 2023. "Education as Opportunity? The Causal Effect of Education on Labor Market Outcomes in Jordan." *Oxford Development Studies* 51, no. 2: 179–197.

Ioannidis, J. P. A., T. D. Stanley, and H. Doucouliagos. 2017. "The Power of Bias in Economics Research." *Economic Journal* 127, no. 605: F236–F265.

Irsova, Z., H. Doucouliagos, T. Havranek, and T. D. Stanley. 2024. "Meta-Analysis of Social Science Research: A Practitioner's Guide." *Journal of Economic Surveys* 38, no. 5: 1547–1566.

Isacsson, G. 1999. "Estimates of the Return to Schooling in Sweden From a Large Sample of Twins." *Labour Economics* 6, no. 4: 471–489.

Isacsson, G. 2004. "Estimating the Economic Return to Educational Levels Using Data on Twins." *Journal of Applied Econometrics* 19, no. 1: 99–119.

Kamhöfer, D. A., and H. Schmitz. 2016. "Reanalyzing Zero Returns to Education in Germany." *Journal of Applied Econometrics* 31: 865–872.

Keane, M., and T. Neal. 2023. "Instrument Strength in IV Estimation and Inference: A Guide to Theory and Practice." *Journal of Econometrics* 235, no. 2: 1625–1653.

Korwatanasakul, U. 2023. "Returns to Schooling in Thailand: Evidence From the 1978 Compulsory Schooling Law." *Developing Economies* 61, no. 1: 3–35.

Kvarven, A., E. Strømland, and M. Johannesson. 2020. "Comparing Meta-Analyses and Preregistered Multiple-Laboratory Replication Projects." *Nature Human Behaviour* 4, no. 4: 423–434.

Lau, J., J. P. A. Ioannidis, N. Terrin, C. H. Schmid, and I. Olkin. 2006. "The Case of the Misleading Funnel Plot." *BMJ* 333, no. 7658: 597–600.

Lee, D. S., J. McCrary, M. J. Moreira, and J. Porter. 2022. "Valid t-Ratio Inference for IV." *American Economic Review* 112, no. 10: 3260–3290.

Leigh, A., and C. Ryan. 2008. "Estimating Returns to Education Using Different Natural Experiment Techniques." *Economics of Education Review* 27: 149–160.

Leon-Bravo, E.. 2022. "Does Compulsory Schooling Impact Labour Market Outcomes? Evidence From the 1993 Educational Reform in Mexico." *Westminster Business School Working Paper Series* 2022/002.

Li, S.. 2023. "Returns to Education for Women in the Mid-Twentieth Century: Evidence From Compulsory Schooling Laws." *Working Paper*, Boston University.

Liu, J. 2024. "Education Legislations That Equalize: A Study of Compulsory Schooling Law Reforms in Post-WWII United States." *Humanities and Social Science Communications* 11: 966. https://doi.org/10.1057/s41599-024-03460-0.

Liwinski, J. 2020a. "The Impact of Compulsory Schooling on Hourly Wage: Evidence From the 1999 Education Reform in Poland." *Evaluation Review* 44, no. 5–6: 437–470.

Liwinski, J. 2020b. "The Impact of Compulsory Education on Employment and Wages in a Transition Economy." *Eastern European Economics* 58, no. 2: 137–173.

Lortie-Forgues, H., and M. Inglis. 2019. "Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?" *Educational Researcher* 48, no. 3: 158–166. https://doi.org/10.3102/0013189X19832850.

Ma, X., and I. Iwasaki. 2021. "Returns to Schooling in China: A Large Meta-Analysis." *Education Economics* 29, no. 4: 379–410.

Meghir, C., and M. Palme. 2005. "Educational Reform, Ability, and Family Background." *American Economic Review* 95, no. 1: 414–424.

Miller, P. W., C. Mulvey, and N. Martin. 1995. "What Do Twins Studies Reveal About the Economic Returns to Education? A Comparison of Australian and U.S. Findings." *American Economic Review* 85, no. 3: 586–599.

Miller, P. W., C. Mulvey, and N. Martin. 2006. "The Return to Schooling: Estimates From a Sample of Young Australian Twins." *Labour Economics* 13, no. 5: 571–587.

Montenegro, C. E. and H. A. Patrinos. 2014. "Comparable Estimates of Returns to Schooling Around the World." World Bank Policy Research Working Paper *No*. 7020.

New, D., C. Sonja, S. Schurer, and D. Sulzmaier. 2021. "Gender Differences in the Lifecycle Benefits of Compulsory Schooling Policies." *European Economic Review* 140: 1–22.

OECD. 2013. *Education at a Glance 2013: OECD Indicators*. OECD Publishing. https://doi.org/10.1787/eag-2013-en.

Oosterbeek, H., and D. Webbink. 2007. "Wage Effects of an Extra Year of Basic Vocational Education." *Economics of Education Review* 26: 408–419.

Oreopoulos, P. 2006a. "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter." *American Economic Review* 96, no. 1: 152–175.

Oreopoulos, P. 2006b. "The Compelling Effects of Compulsory Schooling: Evidence From Canada." *Canadian Journal of Economics* 39, no. 1: 22–52.

Oreopoulos, P., and K. G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25, no. 1: 159–184.

Page, M. J., J. E. McKenzie, P. M. Bossuyt, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *BMJ* 372, no. 71: 1–9.

Parinduri, R. A. 2014. "Do Children Spend Too Much Time in Schools? Evidence From a Longer School Year in Indonesia." *Economics of Education Review* 41: 89–104.

Patrinos, H. A., G. Psacharopoulos, and A. Tansel. 2021. "Private and Social Returns to Investment in Education: The Case of Turkey With Alternative Methods." *Applied Economics* 53, no. 14: 1638–1658.

Patrinos, H. and A. G. Psacharopoulos. 2025. "Causal Returns to Education". GLO Discussion Paper No. 1653.

Pischke, J.-S. 2007. "The Impact of Length of the School Year on Student Performance and Earnings: Evidence From the German Short School Years." *Economic Journal* 117, no. 523: 1216–1242.

Pischke, J.-S., and T. von Wachter. 2008. "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation." *Review of Economics and Statistics* 90, no. 3: 592–598.

Plomin, R. 2011. "Commentary: Why Are Children in the Same Family So Different? Non-Shared Environment Three Decades Later." *International Journal of Epidemiology* 40: 582–592.

Plomin, R. 2024. "Nonshared Environment: Real but Random." *JCPP Advances* 4, no. 1: e12229. https://doi.org/10.1002/jcv2.12229.

Plomin, R. and K. Kawakami. 2023. "Nature, Nurture and Nonshared Environment in Cognitive Development." Working Paper, King's College, University of London.

Powdthavee, N. 2013. “The Marginal Income Effect of Education on Happiness: Estimating the Direct and Indirect Effects of Compulsory Schooling on Well-Being in Australia.” Melbourne Institute Working Paper Series Working Paper No. 16/13.

Purnastuti, L., R. Salim, and M. A. M. Joarder. 2015. “The Returns to Education in Indonesia: Post Reform Estimates.” *Journal of Developing Areas* 49, no. 3: 183–204.

Pustejovsky, J. E. 2017. “You Wanna PEESE of D's?” Working Paper, University of Wisconsin—Madison. Retrieved December 2025. https://jepusto.com/posts/PET-PEESE-performance/#:~:text=Tests%20for%20small%2Dsample%20bias,is%20the%20pooled%20sample%20variance.

Rouse, C. 1999. “Further Estimates of the Economic Return to Schooling From a New Sample of Twins.” *Economics of Education Review* 18, no. 2: 149–157.

Simonsohn, U. 2017. “The Funnel Plot Is Invalid Because of This Crazy Assumption: r(n,d)=0.” Retrieved December 2025. https://datacolada.org/58.

Soon, J.-J., and H.-E. Lim. 2024. “Driver of Returns to Schooling: Education-Related Policies or Family Background?” *GLO Discussion Paper No.* 1471.

Stanley, T. D. 2001. “Wheat From Chaff: Meta-Analysis as Quantitative Literature Review.” *Journal of Economic Perspectives* 15: 131–150.

Stanley, T. D. 2005. “Beyond Publication Bias.” *Journal of Economic Surveys* 19, no. 3: 309–345.

Stanley, T. D., and H. Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business*. Routledge.

Stanley, T. D., and H. Doucouliagos. 2014. “Meta-Regression Approximation to Reduce Publication Selection Bias.” *Research Synthesis Methods* 5: 60–78.

Stephens, M., Jr., and D. Y. Yang. 2014. “Compulsory Education and the Benefits of Schooling.” *American Economic Review* 104, no. 6: 1777–1792.

Torun, H. 2018. “Compulsory Schooling and Early Labor Market Outcomes in a Middle-Income Country.” *Journal of Labor Research* 39: 277–305.

Vieira, J. A. C. 1999. “Returns to Education in Portugal.” *Labour Economics* 6: 535–541.

von Stumm, S., and R. Plomin. 2018. “Monozygotic Twin Differences in School Performance Are Stable and Systematic.” *Developmental Science* 21, no. 6: e12694.

Xue, X., M. Cheng, and W. Zhang. 2021. “Does Education Really Improve Health? A Meta-Analysis.” *Journal of Economic Surveys* 35, no. 1: 71–105.

Zhang, J. 2020. “Estimates of the Returns to Schooling in Taiwan: Evidence From a Regression Discontinuity Design.” *Applied Economics Letters* 27, no. 7: 533–538.

Zhang, J., P.-W. Liu, and L. Yung. 2007. “The Cultural Revolution and Returns to Schooling in China: Estimates Based on Twins.” *Journal of Development Economics* 84, no. 2: 631–639.

Zimmerman, S. D. 2014. “The Returns to College Admission for Academically Marginal Students.” *Journal of Labor Economics* 32, no. 4: 711–754.

Appendix A

Twin Studies

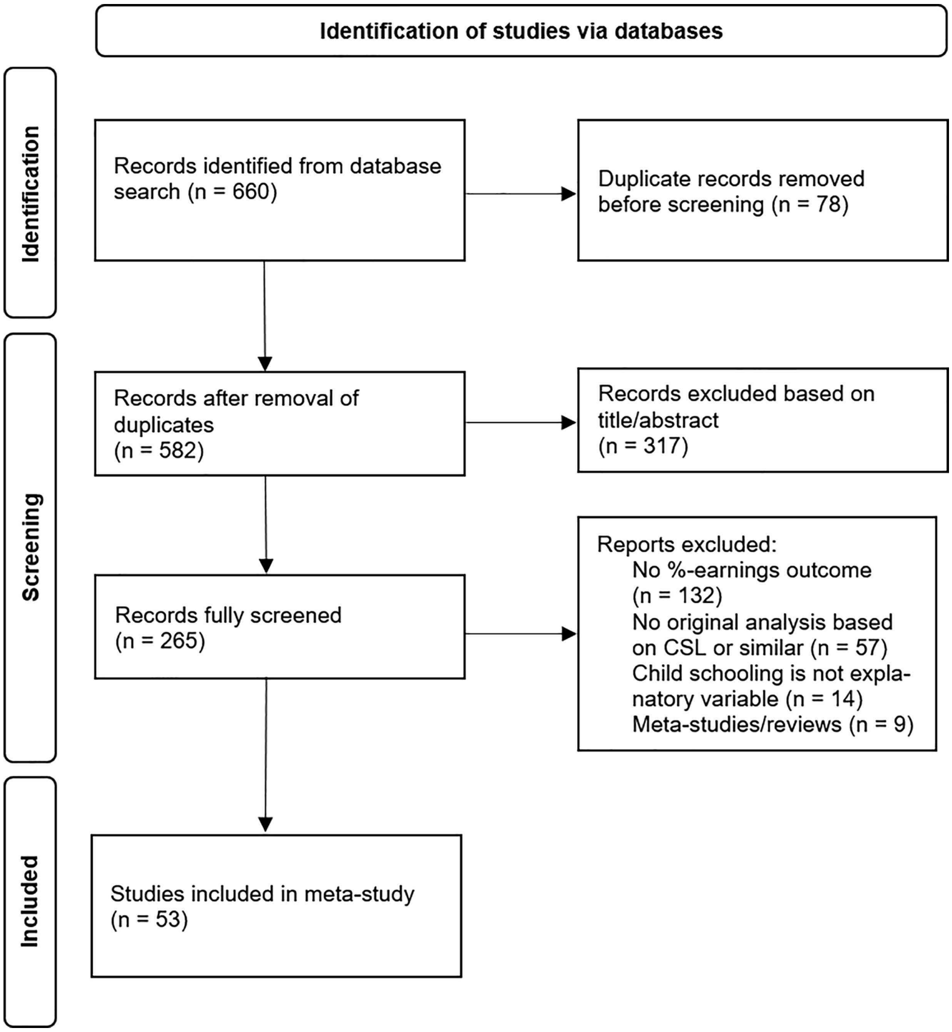
A popular idea, for example, in the 1990s and 2000s, was to use monozygotic (MZ) twins, with the same genetic inheritance, as the ability control in estimating the causal returns to education. See, for example, Amin (2011), Ashenfelter and Krueger (1994), Ashenfelter and Rouse (1998), Behrman and Rosenzweig (1999), Behrman et al. (1994), Bonjour et al. (2003), Isacsson (1999, 2004), Miller et al. (1995, 2006), Rouse (1999), Zhang et al. (2007). Such estimates using the difference in schooling across MZ twins versus their difference in earnings were as high as the raw cross-sectional returns to education, suggesting an absence of significant ability selection into education.

However, these MZ twin estimates do not in fact control for ability. For despite their genetics being the same, monozygotic twins can differ significantly in academic abilities, which helps explain why they sometimes have different years of schooling. Even within the same family environment there is significant variation in MZ twins' phenotypes, despite the same genotype.

On average MZ twins in a British sample, for example, at age 16 differed in academic performance by nearly half a grade level (von Stumm and Plomin 2018, 5). Since these performance differences between the MZ twins at ages 12, 14, and 16 were correlated, this was not just measurement error, but represented persistent academic achievement differences. Robert Plomin, the distinguished behavioral geneticist who has made extensive exploration of the cause of these "nonshared environment" effects, has concluded that whatever these effects are they have no identifiable systematic source (Plomin 2011, 2024; Plomin and Kawakami 2023). But they imply that for MZ twins, as for the general population, differences in schooling are associated with difference in academic abilities.

Appendix B

PRISMA Flow Diagram



Note: Many more than 660 search results appear with search terms such as "returns to compulsory education" on Google Scholar. The specified number is the number of papers we inspected. Few records were removed based on title/abstract relative to other meta-studies because multiple papers with a primary focus on noneconomic outcomes also included estimates of economic returns despite not mentioning it in the title/abstract.

Appendix C

List of Studies in Meta-Analysis

TABLE C1 | The 79 independent estimates in the meta-analysis are taken from the following 53 papers.

Paper	Estimate(s)	Standard error	Status of paper
Angrist and Krueger 1991	8.10 (M)	1.89	Top-5 publication
Harmon and Walker 1995	15.25 (M)	1.50	Top-5 publication
Vieira 1999	0.53 (B)		Publication
Acemoglu and Angrist 2000	9.81 (B)	4.78	Publication
Meghir and Palme 2005	0.88 (M) 2.11 (F)	1.37 1.24	Top-5 publication
Oreopoulos 2006a	15.7 (B) 0.80 (M)	4.40 5.65	Top-5 publication
Oreopoulos 2006b	12.30 (B)	4.63	Publication
Leigh and Ryan 2008	11.80 (B)	3.50	Publication
Oosterbeek and Webbink 2007	−1.80 (B)	1.90	Publication
Pischke 2007	0.70 (B)	1.40	Publication
Pischke and von Wachter 2008	1.60 (B)	1.50	Publication
Brunello et al. 2009	4.89 (M) 6.84 (F)	0.50 0.50	Publication
Aakvik et al. 2010	9.40 (M)	0.20	Publication
Devereux and Hart 2010	3.50 (M) −0.30 (F)	1.70 1.80	Publication
Oreopoulos and Salvanes 2011	13.10 (B)	0.60	Publication
Buscha and Dickson 2012	22.60 (M) 16.90 (F)	21.10 9.00	Publication
Clay et al. 2012	12.60 (M)	8.60	Working paper
Fang et al. 2012	51.00 (M) 10.00 (F)	23.00 5.00	Working paper
Gerritsen 2014	−0.50 (B)	3.6	Working paper
Dickson 2013	10.20 (M)	5.10	Publication
Grenet 2013 (UK)	6.9 (M) 6.4 (F)	2.90 1.10	Publication
Grenet 2013 (France)	1.00 (M) −0.40 (F)	3.00 1.30	Publication
Powdthavee 2013	15.50 (M) 10.90 (F)	3.30 4.30	Working paper
Parinduri 2014	14.00 (M) 26.00 (F)	8.00 10.00	Publication
Stephens and Yang 2014	−0.30 (B) −1.40 (M)		Publication
Alzúa et al. 2015	16.77 (B) 22.77 (M)	4.83 5.73	Publication

(Continues)

TABLE C1 | (Continued)

Paper	Estimate(s)	Standard error	Status of paper
Buscha and Dickson 2015	−1.00 (B)	0.56	Working paper
Purnastuti et al. 2015	22.1 (M) −6.4 (F)	17.7 0.65	Working paper
Chib and Jacobi 2016	1.15 (B)	7.4	Publication
Kamhöfer and Schmitz 2016	0.01 (B)	3.36	Publication
Aydemir and Kirdar 2017	2.00 (M) 5.70 (F)	0.80 0.26	Publication
Bhuller et al. 2017	11.20 (B)	4.80	Publication
Dolton and Sandi 2017	5.70 (M)	3.14	Publication
Torun 2018	1.70 (M) 9.80 (F)	1.00 1.70	Publication
Delaney and Devereux 2019	6.00 (B)	1.18	Publication
Eble and Feng 2019	0.20 (M) 4.00 (F)	1.10 1.10	Publication
Albarrán et al. 2020	17.60 (M) 22.40 (F)	9.00 8.50	Publication
Chen et al. 2020	12.70 (B)	7.70	Publication
Fischer et al. 2020	1.80 (M) 3.50 (F)	0.50 1.70	Publication
Liwinski 2020a	12.00 (M) 7.40 (F)	6.20 10.70	Publication
Liwinski 2020b	0.50 (M) −6.20 (F)	2.90 13.3	Publication
Zhang 2020	4.92 (B)	0.20	Publication
Clay et al. 2021	7.70 (M)	1.50	Publication
de New et al. 2021	0.00 (B)		Publication
Domnisoru 2021	3.30 (M) 1.40 (F)	1.30 1.10	Publication
Elsayed and Marie 2020	13.50 (M) 28.30 (M)	7.20 13.2	Working paper
Patrinos et al. 2021	18.00 (M) 11.20 (F)		Publication
Fischer et al. 2022	5.9 (B)		Publication
Leon-Bravo 2022	7.33 (B)	7.10	Working paper
Clark 2023	1.21 (M) 1.43 (F)	1.35 1.70	Publication
Hicks and Duan 2023	5.10 (B)		Publication
Korwatanasakul 2023	7.90 (M) 8.31 (F)	0.93 0.72	Publication
Li 2023	15.10 (F)	2.80	Working paper
Liu 2024	7.40 (B)	0.50	Publication

Note: All estimates in percentages. B = both genders, F = female, M = male. For more information on where estimates from specific papers are taken from, see the Excel or dta. file in the online appendix ("source_effect" variable).

Appendix D

List of Reforms in Meta-Analysis

Nr.	Reform	Paper(s)
1	Argentina, MYOS extension, 7 to 9 years, in 1993	Alzúa et al. 2015
2	Australia, MSLA extension, 14 to 15 years, in 1945 (adoption until 1960/1963)	New et al. 2021
3	Australia, multiple across state variations and changes in CSLs	Leigh and Ryan 2008, Powdthavee 2013
4	Austria, MSLA extension, 14 to 15 years, in 1962/66	Brunello et al. 2009, Albarrán et al. 2020
5	Belgium, MSLA extension, 14 to 18 years, in 1983	Brunello et al. 2009
6	Canada, multiple across state variations and changes in CSLs	Oreopoulos 2006b
7	China, MYOS extension, 5 to 6 years, in late 1970s	Eble and Feng 2019
8	China, secondary school decrease from 3 to 2 back to 3 years, 1981–1985 (staggered rollout)	Chen et al. 2020
9	China, 9 year compulsory education established, in 1986 (staggered rollout)	Fang et al. 2012
10	Czechia/Slovakia, MYOS, 8 to 9 years, in 1960	Albarrán et al. 2020
11	Denmark, MYOS extension, 7 to 9 years, in 1971	Albarrán et al. 2020
12	Egypt, MYOS decrease, 9 to 8 years, 1989	Elsayed and Marie 2020
13	Finland, MSLA extension, 13 to 16 years, in 1972–1977 (staggered rollout)	Brunello et al. 2009
14	France, MSLA extension, 15 to 16 years, in 1967	Grenet 2013, Albarrán et al. 2020, Domnisoru 2021
15	Germany, MYOS extension, 8 to 9 years, in 1949–1970 (staggered rollout across states)	Pischke and von Wachter 2008, Brunello et al. 2009, Kamhöfer and Schmitz 2016
16	Germany, reform change leading to shorter school year, in 1966–1967	Pischke 2007
17	Greece, MSLA extension, 12 to 15 years, in 1972	Brunello et al. 2009, Albarrán et al. 2020
18	Hungary, MYOS extension, 8 to 10 years, in 1961	Albarrán et al. 2020
19	Indonesia, school term length increase, in 1978–1979	Parinduri 2014
20	Indonesia, MYOS extension, 6 years, in 1984	Purnastuti et al. 2015
21	Indonesia, MYOS extension, 6 to 9 years, in 1994	Purnastuti et al. 2015
22	Ireland, MSLA extension, 14 to 15 years, in 1972	Brunello et al. 2009
23	Italy, MSLA extension, 11 to 14 years, in 1963	Brunello et al. 2009, Albarrán et al. 2020
24	Jordan, MYOS extension, 9 to 10 years (also reconstruction of secondary schooling), in 1998	Hicks and Duan 2023
25	Malta, MYOS extension, 8 to 10 years, in 1972	Albarrán et al. 2020
26	Mexico, MSLA, 12 to 15 years, 1993	Leon-Bravo 2022
27	Netherlands, MSLA extension, 14 to 15, in 1971	Gerritsen 2014
28	Netherlands, MSLA extension, 15 to 16 years, in 1975	Brunello et al. 2009, Oosterbeek and Webbink 2007, Albarrán et al. 2020
29	Northern Ireland, MSLA extension, 14 to 15 years, in 1957 (related to 1947 United Kingdom reform)	Oreopoulos 2006a, Devereux and Hart 2010
30	Norway, MYOS extension, 7 to 9 years, in 1960–1975, also change of curriculum, etc.	Aakvik et al. 2010, Bhuller et al. 2017
31	Poland, MYOS extension, 7 to 8 years, in 1966	Albarrán et al. 2020, Liwinski 2020b
32	Poland, MYOS extension, 8 to 9 years, in 1999	Liwinski 2020a
33	Portugal, MSLA extension, 11 to 12 years, in 1956	Vieira 1999

Nr.	Reform	Paper(s)
34	Portugal, MSLA extension, 12 to 14 years, in 1964	Vieira 1999, Albarrán et al. 2020
35	Spain, MSLA extension, 12 to 14 years, in 1970	Brunello et al. 2009, Albarrán et al. 2020
36	Sweden, extension of term length from 34.5/36.5 to 39 weeks in 1939	Fischer et al. 2020
37	Sweden, broad reform including MYOS extension (plus abolishing tracking), 8 to 9 years, in 1948–1953 (staggered rollout)	Meghir and Palme 2005, Fischer et al. 2020, Fischer et al. 2022,
38	Sweden, MYOS extension, 7 to 8 years, in 1941–1962 (staggered rollout)	Brunello et al. 2009
39	Taiwan, MYOS extension, 6 to 9 years, in 1968	Zhang 2020
40	Thailand, MYOS extension, 4 to 6 years, in 1978	Korwatanasakul 2023
41	Turkey, MYOS extension, 5 to 8 years, in 1997	Aydemir and Kirdar 2017, Torun 2018, Patrinos et al. 2021
42	United Kingdom, MSLA extension, 14 to 15 years, in 1947	Harmon and Walker 1995, Oreopoulos 2006a, Devereux and Hart 2010, Chib and Jacobi 2016, Dolton and Sandi 2017, Clark 2023
43	United Kingdom, reform change leading to children born in certain months to have more compulsory schooling and other changes, 1963	Dolton and Sandi 2017
44	United Kingdom, MSLA extension, 15 to 16 years, in 1972	Buscha and Dickson 2012, Dickson 2013, Buscha and Dickson 2015, Dolton and Sandi 2017, Delaney and Devereux 2019, Albarrán et al. 2020
45	United States, multiple across state variations and changes in CSLs	Angrist and Krueger 1991, Acemoglu and Angrist 2000, Oreopoulos and Salvanes 2011, Clay et al. 2012, Stephens and Yang 2014, Clay et al. 2021, Li 2023
46	United States, MSLA extension, 14 to 16 years, multiple Southern states, 1946	Liu 2024

Note: The specified date is the year of implementation of the reform if reported (otherwise the year of the reform decision). Papers report either MSLA = minimum school-leaving age or MYOS = minimum years of schooling.

Appendix E

Summary Statistics and Codebook for Other Variables

Variable name	Description	Mean	Standard deviation	*n*
effect_ percentage	Estimated effect size/return to an extra year of schooling in percentage	8.23	8.83	79
se_ percentage	Standard error of estimates	4.39	4.70	71
sig	Binary indicator of whether estimate is significant at 5% (=1) or not (=0)	0.52	0.50	79
tstat	*t*-statistics of estimate	327.95	2749.37	72
fstat	*f*-statistics of first stage estimation	210.03	586.55	49
n	Number of observations	517158.20	1,812,123	63
full	Binary indicator of whether the estimate is from a full sample (=1) or not (=0)	0.30	0.46	79
men	Binary indicator of whether the estimate is from a male sample (=1) or not (=0)	0.41	0.49	79
women	Binary indicator of whether the estimate is from a female sample (=1) or not (=0)	0.29	0.46	79
year_of_reform	The year the reform is implemented	1967.76	20.36	79
year_since_pub	Years since paper was published (as of 2024)	9.30	6.81	79

Variable name	Description	Mean	Standard deviation	*n*
wp	Binary indicator of whether the paper is a working paper (=1) or published (=0)	0.14	0.35	79
scholar_ citations	Number of citations on Google Scholar as of September 2024	233.80	530.03	79
jour_imp_ factor_2022	The journal impact factor in 2022 of the journal where the paper is published	3.76	3.34	68
ols	Binary indicator of whether the paper reports OLS estimate(s) (that is comparable to IV estimate(s))	0.74	0.44	58
ols_effect	Uncontrolled OLS correlation between years of schooling and earnings	7.04	3.17	43
iv_biggest	Binary indicator of whether IV produces the biggest estimate (=1) or OLS produces the biggest estimate (=0)	0.60	0.49	43
poor-middle	Binary indicator of whether a country has above (=0) or below (=1) 10.000 USD GDP per capita in the year 2000	0.34	0.45	79
gdp_p_c	GDP per capita in 2000 in current USD	19175.97	13106.7	75
MSLA before reform	Minimum school-leaving age before the reform was implemented	13.40	1.43	60

Note: Means and standard deviations are taken for the sample of 79 independent estimates. However, many papers have some missing information explaining fewer observations for some variables. If an interval is given for year of reform, the median year is reported.

Other coded variables.

Variable name	Description
Paper	Name of author(s) and year of publication
country	Which country (ies) the paper is researching
csl_change	Description of change in compulsory schooling law that is researched
outcome	What outcome variable the paper is researching
data_type	Type of data (cross section, repeated cross section, and panel)
method_estimation	Estimation method (DD = difference in differences, IV = instrumental variable, RDD = regression discontinuity design)
ind_est	Binary indicator of whether the estimate is independent (=1) or not (=0)
subgroup	Binary indicator of whether the paper reports in subgroups (=1) or not (=0)
subgroup_only	Binary indicator of whether the paper **only** reports in subgroups (=1) or not (=0)
claim_sig	Binary indicator of whether authors claim significant effects of education on their outcome in abstract, introduction and/or conclusion
actual_sig	Binary indicator of whether preferred causal estimates (or most of reported causal estimates in cases without preferred model) are significant at 5%
source_effect	Description of how the estimate was found/summarized
source_n	Description of how the sample size was found/summarized
source_fstat	Description of how *f*-statistics was found/summarized
any_sig	Binary indicator of whether there are **any** significant estimates (either men, women, full sample, or multiple of them)
new_paper_indicator	Binary indicator indicating a new paper
pioneer	Binary indicator of the pioneering study (=1) and the rest (=0)
comments	Comments when considered necessary to avoid confusion and explain decisions

Note: Means and standard deviations excluded for these variables since they either bring no meaningful information or are not actively used in the analysis.

Appendix F

Funnel Plot

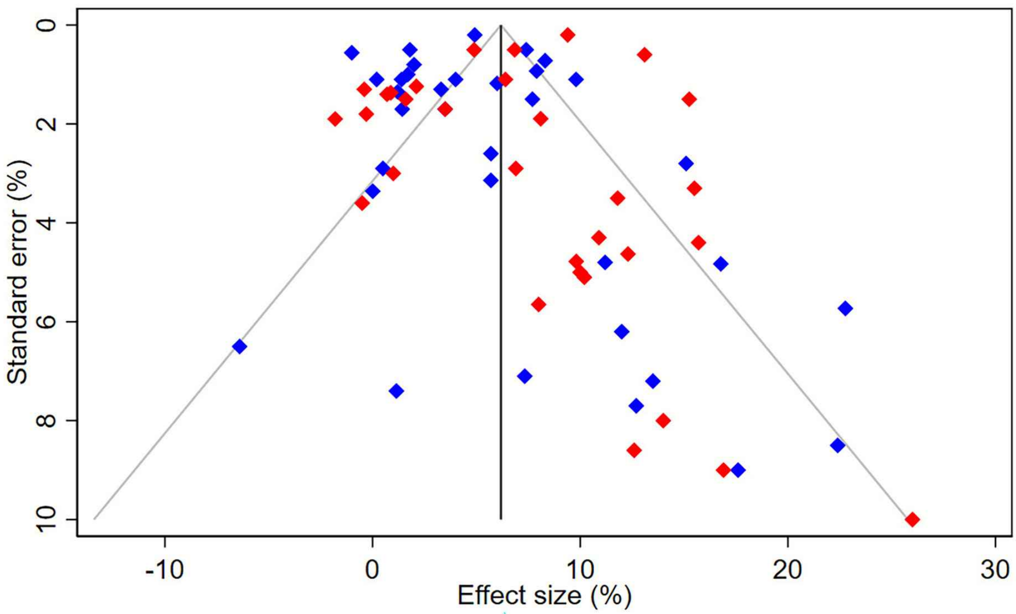


FIGURE F1 | **Funnel plot of rate of return and standard error.** *Note:* $N = 65$. Six outliers with a standard error above 10% omitted. Red diamonds are estimates from old papers (1991–2014), while blue diamonds are estimates from new papers (2015–2024). Funnel plot based on Egger et al. 1997 that weights estimates by their precision. *Source:* Meta-study dataset in online appendix. [Colour figure can be viewed at wileyonlinelibrary.com]