The Labor Market Outcomes of Formal Vocational Education and Training: A Meta- Analysis

Working Paper

Author(s):
Bolli-Kemper, Johanna Mirka; Au Yong Lyn, Audrey; McDonald, Patrick; Renold, Ursula

Publication date:
2024-05

Permanent link:
https://doi.org/10.3929/ethz-b-000675898

Originally published in:
CES Working Paper Series 19

ETH zürich

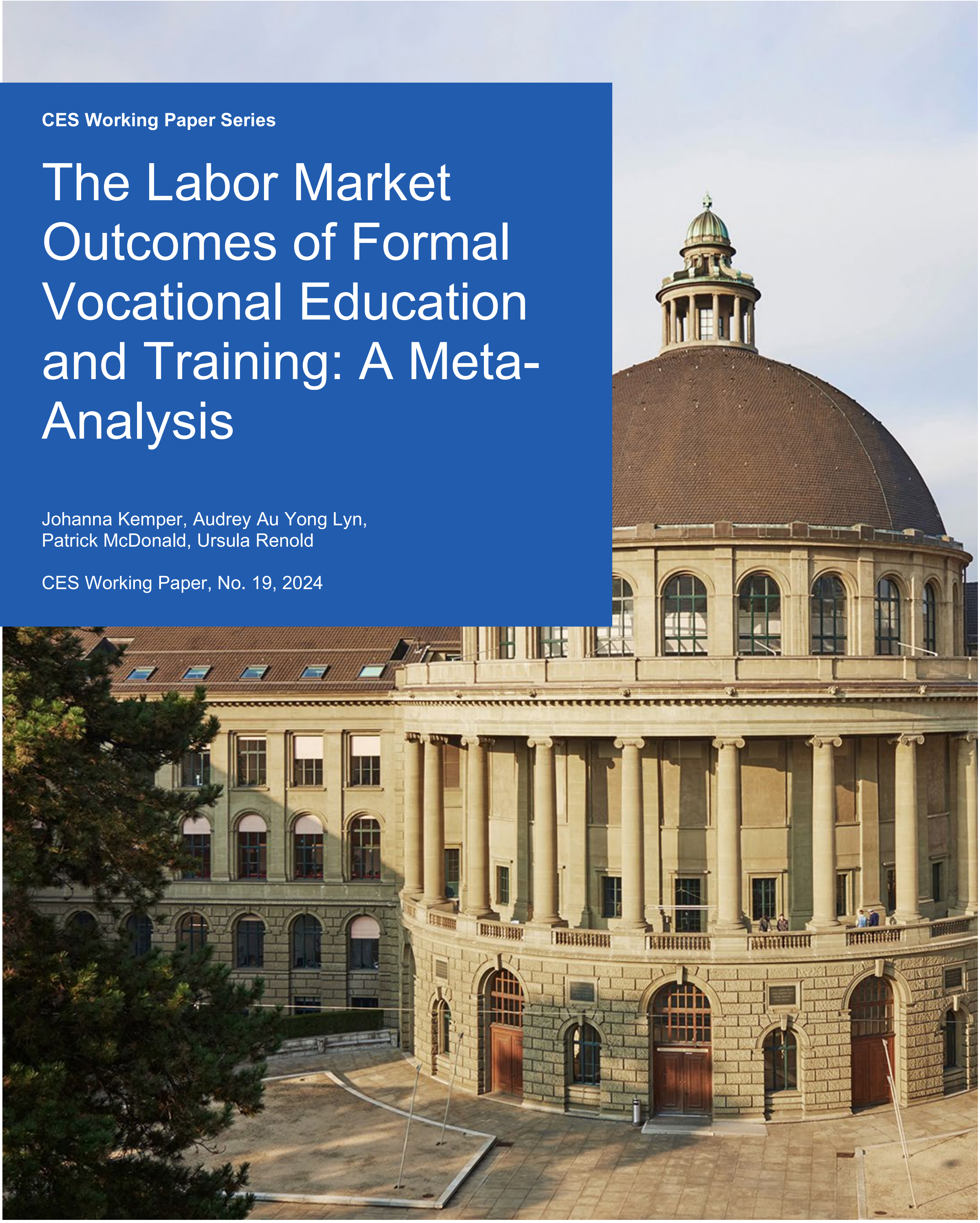
CHAIR OF EDUCATION SYSTEMS

CES Working Paper Series

The Labor Market Outcomes of Formal Vocational Education and Training: A Meta-Analysis

Johanna Kemper, Audrey Au Yong Lyn,
Patrick McDonald, Ursula Renold

CES Working Paper, No. 19, 2024



D MTEC

The Labor Market Outcomes of Formal Vocational Education and Training: A Meta-Analysis

Johanna Kemper[1], Audrey Au Yong Lyn[1,2], Patrick McDonald[1], Ursula Renold[1]

Abstract

This meta-analysis empirically synthesizes the impact of formal Vocational Education and Training (VET) on labor market outcomes using studies published between 1990 and 2022, employing a causal counterfactual evaluation design. To summarize effect sizes across studies, we employ the standardized mean difference Hedges' g for all outcomes, and employment probability and earnings/wages for a subset of studies. We use Robust Variance Estimation (RVE) to account for dependent effect sizes within and between studies. Based on a sample of 39 studies, the impact of VET versus general education on all labor market outcomes is positive but small (Hedges' g = 0.021***) and tends to zero over time, while its impact on employment is minor (4.8%***) and not statistically significant for earnings/wages. Studies with more rigorous evaluation designs report lower impacts. Conversely, based on a sample of 27 studies, the impact of VET versus an unrestricted control group is medium (Hedges' g = 0.071***), depicting significant boosts in employment (11.7%***) and earnings/wages (37.6%***).

[1]Chair of Education Systems (CES), ETH Zurich, Stampfenbachstrasse 69, Zurich, 8092, Switzerland.
[2]Corresponding author: Audrey.AuYongLyn@mtec.ethz.ch

1. Introduction

Formal Vocational Education and Training (VET) is an important component of the education systems in many OECD countries, where about 42% of all pupils at the upper-secondary education level were enrolled (OECDa, 2023). VET plays a crucial role in engaging students and facilitating their transition from school to work. The evolving work environment underscores the importance of VET in developing skills that align with the needs of today's labor markets and societies (OECDb, 2023).

So far, there are only descriptive summaries of empirical studies (Dougherty & Ecton, 2021; Carruthers & Jepsen, 2021; McNally et al., 2022) to inform policy makers about the impact of *formal* VET on labor market outcomes. Formal VET implies that programs must be part of the formal education system of a given country, i.e., must be formally recognized by the Ministry of Education. This explicitly excludes nonformal vocational training that is mostly delivered in the form of Active Labor Market Programs (ALMPs), which has already been evaluated by several meta-analyses (e.g., Kluve et al., (2019), Kemper et al., (2022)). Subsequently, this is the first meta-analysis to empirically summarize the impact of *formal* VET, undertaken at the secondary, post-secondary or tertiary level of the education system, on labor market outcomes of studies published 1990-2022 employing experimental or quasi-experimental counterfactual evaluation designs, i.e., using treatment and control-group design. Our meta-analysis covers a large sample of 70 studies. We code a large variety of program, participant, and study characteristics to account for potentially heterogenous impacts of VET.

Through our extensive online search for relevant studies, which we conducted in April 2021 and October-November 2022, we found 70 studies belonging to three different strands of the literature evaluating the impact of VET on labor market outcomes. Namely, 39 studies analyzing the impact of VET versus general education, 27 studies examining its impact versus an unrestricted control group that is not constrained in their outside option, and four studies looking at the impact of work- versus school-based VET. Given the low number of four studies belonging to the third strand, we restrict our empirical analysis to the first two strands of the literature.

To summarize effect sizes across primary studies using different outcomes measures and different scales, we compute the standardized mean difference (SMD) Hedges' g (Hedges, 1981). However, since Hedges' g is a unitless measure, its magnitude has no economic interpretation and must be compared to field-specific benchmarks (Kraft, 2020). Therefore, we display the economic impact of VET on employment probability and earnings/wages for a subset of studies.

Dependencies of effects sizes within and between studies are a common issue in meta-analysis, which mostly arise if one study reports multiple outcomes, or two or more studies use the same dataset. To account for statistically dependent effect sizes, we use Robust Variance Estimation (RVE) to synthesize effect sizes (Hedges et al., 2010).[3] To test the robustness of our results, we employ Unrestricted Weighted Least Squares (UWLS) as suggested by (Stanley & Doucouliagos, 2017) and Ordinary Least Squares (OLS).

[3]In the past, RVE was more popular in the field of psychology and health, but more recently also in economics (Balasubramanian et al., 2024; Kemper et al., 2022; Waheed, 2023; Kaiser & Menkhoff, 2020; Kaiser et al., 2022).

One large strand of the literature evaluating the impact of VET on labor market outcomes analyzes the effect of VET versus general education. A dominant view of this literature argues that there is a clear trade-off between early career benefits and long-term adaptability of VET. Accordingly, VET tends to provide specific skills that facilitate a smoother transition from school to work, resulting in initially relatively better labor market outcomes. Further, this advantage would diminish over time as the specific skills acquired in VET programs become obsolete relatively more quickly (Hampf et al., 2017). However, consistent with existing research employing experimental or quasi-experimental evaluation designs, results of our meta-analysis show that the net effect of VET versus general education is positive albeit marginal (Hedges' g= 0.021***) and tends to zero over time. The economic impact on employment is also minor (4.8%***) and lacks statistical significance in terms of earnings/wages. Studies employing more rigorous counterfactual evaluation designs report comparatively lower average impacts.

Another large strand of the literature evaluates the impact of VET versus an unrestricted control group that is not constrained in their outside option. This strand mostly finds positive effects of VET (Dynarski et al., 2018). Our results suggest that the impact of VET is generally positive, and to be considered as medium (Hedges' g=0.071***) when compared to benchmarks specific to the field of education (Kraft, 2020). A subset of these studies indicates that VET boosts employment by 11.7%*** and earnings/wages by 37.6%*** relative to the control group. Approximately two-thirds of these studies focus on Associate Degree programs at Community Colleges in the US, which demonstrate larger average effect sizes compared to other studies. Albeit being based on a small number of studies, an analysis of the impact by field of Associate Degree programs shows the highest returns in Health Science and Agriculture, with earnings/wages 54.7%*** and 24.7%*** higher than those of the control group, respectively.

The paper is organized as follows. Section two specifies which studies we include and how they were searched for and coded. The third section describes the outcomes, measures and empirical methods used for our meta-regressions. Section four provides the results and section five concludes.

2. Sample construction

2.1 Inclusion criteria

We searched for studies published between 1990 and 2022 evaluating the impact of *formal* VET programs on labor market outcomes of persons of any age.[4] We define a "formal" VET program as being part of the formal education system of a given country. This is fulfilled if a VET program is recognized and mainly run by the Ministry of Education (UNESCO Institute for Statistics., 2012).[5] We include programs that deliver VET in classrooms only, or as a combination of classroom- and workplace-based

[4] We do not select our sample of studies based on a certain age threshold to make sure we include all relevant studies.

[5]The Ministry of Labor may be also involved, especially in programs that involve workplace-based education, but is not the main actor.

education.[6][7] The programs can be part of the formal secondary, post-secondary or tertiary education level. [8]

We include studies that employ a causal counterfactual evaluation design; that is, studies evaluating the impact of VET employing a treatment and control group. We distinguish between three different categories of such evaluation designs (Imbens, 2023). First, *experimental designs* (e.g., Randomized Controlled Trials, RCTs), where individuals are randomly assigned to either treatment or control group. Thereby, treatment is said to be statistically independent of potential outcomes (Wooldridge, 2010). Second, evaluation designs *with unconfoundedness*, where treatment is only statistically independent of potential outcomes conditional on covariates (Imbens & Wooldridge, 2007). Hence, regression methods are used to adjust for observed confounders. In our meta-analysis, this includes matching designs. Third, evaluation designs that allow for credible estimation of causal effects *without* the *unconfoundedness* assumption (Imbens, 2023).[9] In our meta-analysis, this includes instrumental variable estimation (IV), fixed effect (FE) and difference in differences methods (DiD), and regression discontinuity designs (RDDs). Among these three categories, experimental evaluation ("experimental") designs are considered to have to have the lowest risk-of-bias, followed by designs *without unconfoundedness* ("without unconfound.") assumption. Designs *with unconfoundedness* ("with unconfound.") assumption are considered to have the highest risk-of-bias as they are the least restrictive in terms of causal identification.

In terms of geographic location, we add studies focusing on High-income countries (HICs) or Low-/Middle-income countries (L/MICs) published in peer-reviewed journals, working papers or reports. We ensure that these studies estimate at least one of the following primary labor market outcomes: employment probability, labor force participation, hours worked, earnings or wages.[10]

The literature on the impact of VET on labor market outcomes can be divided in three strands. Studies belonging to the first stand of the literature evaluate the impact of VET versus general education. As such, these compare the effect of VET versus a more restricted control group than studies belonging to the second strand of the literature, where the control group may pursue any kind of outside activity (e.g., enter labor market, further education). Studies belonging to the third strand of the literature analyze an even more restricted treatment and control group by evaluating the impact of programs that combine workplace- and classroom -based VET versus fully classroom -based VET programs.

2.2 Search process

Our strategy for finding suitable studies consisted of three main steps. First, we screened general (e.g. Google Scholar) and specialized databases (e.g. IDEAS/RePEc), searched through forward citations based on key papers using Google Scholar, and by screening reference lists in April 2021 and October-

[6] Thereby, the workplace-based component must be part of the formal curriculum. We do not count voluntary internships or the like as workplace-based education.
[7] This includes dual apprenticeships as present in the German-speaking countries (Ryan, 2012).
[8] We do not explicitly search for continuing education and training programs, since these serve a different target a population with different characteristics, e.g., different levels of work experience, hence productivity.
[9] According to (Imbens, 2023), in such cases, additional variables with specific causal structure, or additional assumptions placed on either the assignment mechanism or the potential outcome distributions then help to identify treatment effects.
[10] The difference between earnings and wages is that wages are conditional on being employed.

November 2022. Second, we contacted authors of US studies that we chose to be included in our meta-analysis asking if they knew more relevant studies than those we provided them in a reference list. This subsequently resulted in 7237 studies, including 848 papers we found as a byproduct of an extensive online search for a meta-analysis evaluating the labor market impact of vocational training in the context of active labor market programs (ALMPs) (Kemper et al., 2022). After removing duplicates and screening the title and abstract of studies, we examined the remaining 251 studies based on their full text. This process led to a final sample of 70 studies (see Figure A1).

2.3 Computing estimates and information extraction

To code studies, we used a coding sheet and a manual provided Kluve et al. (2019). This ensured a consistent method accounting for various studies, programs, and participant characteristics, that had been previously in other peer-reviewed publications. Each study was completely coded by one person and a second person reviewed the coding. If information on program characteristics was not available from the study, coders conducted additional online research.[11]

3. Methodology

3.1 Outcome measures

3.1.1 Standardized mean difference (SMD) Hedges' g

It is common for studies to use different outcome measures to quantify the impact of VET on labor market outcomes. However, different units of measurement (e.g. logarithm of wages and wages in terms of US$) are often used across various studies, making it difficult to summarize respective outcome measures to one. One way to circumvent this problem is to use an outcome measure that makes such effect sizes comparable. We do so by employing the standardized mean difference (SMD) Hedges' g (Hedges, 1981), which summarizes effect size estimates of the included studies in a unitless manner. The magnitude of Hedges' g itself has no economic interpretation. It is computed as:

$$g = \frac{\overline{Y}_t - \overline{Y}_c}{SD_p} \times \left(1 - \left(\frac{3}{4 \times (n_t + n_c) - 9}\right)\right), \qquad (1)$$

where $\overline{Y}_t$ and $\overline{Y}_c$ represent the mean outcomes and n_t and n_c the sample sizes in treatment and control groups. Their difference captures the treatment effect of vocational education. To obtain a standardized measure, the treatment effect is divided by the pooled standard deviation SD_p.[12]

Accordingly, the standard error of Hedges' g is given by:

$$SE_g = \sqrt{\left[\frac{n_t + n_c}{n_c * n_t} + \frac{g^2}{2 * (n_c + n_t)}\right]} \qquad (2)$$

Hedges' g requires information about the mean and standard deviation of the outcome variable for treatment, control and pooled sample. Since only a few papers report all these metrics, we approximate

[11] Coding sheets, manual and additional resources can be made available upon request.

[12] More precisely, $SD_p = \sqrt{\frac{(n_c - 1) \times SD_c^2 + (n_t - 1) \times SD_t^2}{n_c + n_t - 2}}$ where SD_t and SD_c represent the sample standard deviations of outcomes in the treatment and control group.

the pooled standard deviation using the Borenstein (2009) formula $SD_p \approx SE \times \sqrt{\frac{n_t \times n_c}{n_t + n_c}}$, where SE is the standard error of the treatment effect.

While Hedges' g can combine various outcome measures with different scales, it lacks a direct link to the economic significance of effects. Commonly, Standard Mean Differences (SMD) are measured against established standards like those of Cohen (2013), who categorized SMDs of 0.2 as small, 0.5 as medium, and beyond 0.8 as large. However, Cohen's standards stem from a few small, tightly controlled social psychology lab experiments conducted in the 1960s, primarily involving undergraduates (Kraft, 2020). These criteria might not apply uniformly across all research areas. In the context of field-based educational research, Kraft (2020) suggested new empirical standards for evaluating effect sizes. Kraft (2020) builds on nearly 750 RCTs measuring impacts of education interventions on academic achievement. Accordingly, in social sciences, it is more appropriate to view effect sizes measured in terms of their standard deviations, just as the SMD, below 0.05 as minor, between 0.05 and 0.20 as moderate, and above 0.20 as significant. In the following, we therefore compared the magnitude of our SMDs with the benchmark of Kraft (2020).

3.1.2 Employment probability and earnings/wages

To quantify the economic effect of VET, we analyze its impact on employment probability and the logarithm of earnings/wages. Although information on these labor market outcomes exists for about half of our studies, they should be sufficient to derive empirically reasonable estimates of the effect of formal VET on employment and wages.

3.2 Robust variance estimation

Dependence between effect size parameters and error terms is a common problem in meta-analysis, as they often report multiple effect sizes per study. This may occur if the same individual is measured by different outcomes or at different points in time, or because different individuals are compared to the same control group. Such dependency violates the core assumption of meta-analysis: that effect sizes are independent (Hedges et al., 2010). Ignoring this dependence can lead to downward-biased standard errors (López-López et al., 2017). Though the exact dependence structure between effect sizes is almost never fully known, having such information would allow specifying weights that are exact inverse variance to obtain the most efficient estimator of the weighted mean, standard errors, test statistics, and confidence intervals.

Historically, when faced with multiple, dependent effect sizes, the common practice was to choose just one effect size per study or to compute an average of the effect sizes at the study level, as noted by Borenstein (2009). However, this tends to inflate standard errors and disregards information varying at the study level (López-López et al., 2017). Another approach models the dependency structure by means of multivariate meta-regression requiring information on the true dependency structure of effect sizes, which is however rarely reported (Hedges et al., 2010). Another alternative is multi-level meta-analysis (MLMA) assuming that effect sizes within studies are independent, which is violated when primary studies report multiple effect sizes based on the same sample (Cheung, 2019).

To account for effect size dependency in our study, we subsequently employ robust variance estimation (RVE) which can circumvent the aforementioned concerns. RVE uses a Weighted Least Squares (WLS) estimator with weighting matrices specifying the approximate dependence structure between effect sizes by using the cross-products of the regression residuals as a crude estimate of the covariances between effect size estimates. This estimate has been established to be accurate for a sample of more than 40 studies (Pustejovsky & Tipton, 2022; Hedges et al., 2010). For samples with 40 studies or less, small-sample adjustment methods are available (Tipton E. , 2015).

To estimate RVE models, three critical choices must be made. First, RVE requires determining the unit of analysis (or cluster) where most of the correlation between effect sizes at the study level takes place. RVE assumes that effect sizes across these clusters are independent, while effect size estimates within clusters can be correlated (Hedges et al., 2010). Studies included in our meta-analysis mostly report multiple outcome measures for the same cohort of VET participants. With a few exceptions, most studies evaluate different VET programs for different cohorts rather than the same VET program for the same cohort. Hence, most of the dependency between effect sizes in our study occurs at the level of a cohort participating in a given VET program. We conclude that the "cohort-by-program cluster" is the most appropriate for analysis. In the following, we refer to this simply as "cluster". Second, choosing the unit of analysis directly affects which 'working' model is the most appropriate for our analysis: either the correlated effects or the hierarchical effects model (Hedges et al., 2010).[13] The correlated effects model is chosen, under the assumption that dependencies primarily occur within the same cluster, such as multiple outcome measures in one study, including only a between-study random effects component. The hierarchical model, in contrast, suggests dependencies arise from clusters nested within larger clusters, like multiple studies on the same intervention. As can be seen from the argumentation for the choice of unit of analysis, the correlated effects model is the better fit in the context of our study. Third, RVE estimation involves choosing weighting based on either a fixed- or random-effects model (Tipton E. , 2013). Considering the diversity of our sample in terms of intervention features, geographic locations, methodologies, and outcomes, we opt for a random-effects model.

Our final RVE model, using the correlated-effects model, explicitly addresses both the heterogeneity between clusters and the measurement errors within clusters. The model is formulated according to the following specification:

$$g_{ij} = \alpha + X'_{ij}\beta + v_j + \varepsilon_{ij} \tag{3}$$

Here, g_{ij} represents the *i*-th effect size estimate within the *j*-th cluster. The term α represents the mean of the distribution of true effects across clusters, X'_{ij} is a vector of covariates included in our multivariate meta-regressions.[14] $v_j \sim N(0, \tau^2)$ is the cluster level random effect and $\varepsilon_{ij} \sim N(0, \sigma^2)$ is the residual of the *i*-th effect size estimate within cluster *j*. Moreover, $Var(v_j) = \tau^2$ represents the between-cluster variance component in true effects, which is unknown and needs to be estimated from the data.

[13] More recently, Pustejovsky & Tipton (2022) advocate a working model combining correlated and hierarchical effects.

[14] Note that we do not centre our explanatory variables. This implies that we do not estimate the within- and between-cluster effect of covariates separately but instead pool both effects together in our multivariate regressions in accordance with Tanner-Smith and Tipton (2014).

Estimation of τ^2 requires to make an assumption about the within-cluster correlation in effect sizes, which we set at the default value of σ= 0.8 (Hedges et al., 2010).[15] The estimated $\hat{\tau}^2$ then informs the calculation of inverse variance weights, given as $w_{ij}^{RVE} = \frac{1}{n_j(\sigma_j^2)+\hat{\tau}^2}$, with n_j as the number of estimates and σ_j^2 as the average sampling variance within cluster *j*.

3.3 Robustness

In order to test the robustness of our results, we use two alternative estimation methods: unrestricted weighted least squares (UWLS) as described by Stanley & Doucouliagos (2017), Stanley et al. (2022) and ordinary least squares (OLS).

3.3.1 Unrestricted Weighted Least Squares (UWLS)

UWLS assumes that between-study heterogeneity is proportional to the within-study sampling error (Stanley et al., 2022). As a result, weights in UWLS are assigned as the inverse variances of the effect sizes: $w_{ij}^{UWLS} = \frac{1}{\gamma\sigma_{ij}^2}$ in the following model specification:

$$g_{ij} = \alpha + X'_{ij}\beta + \varepsilon_{ij} \qquad (4)$$

UWLS is conceptually positioned between the traditional common- and random-effects models, which means that the average effect estimates from UWLS are the same as those from the common-effects model, but with broader confidence intervals, especially when there is significant variance between studies (Borenstein et al., 2010).

Although Robust Variance Estimation (RVE) also uses inverse variance weights, it differs from UWLS in three main ways. First, UWLS gives more weight to more precise estimates and studies with multiple effect sizes by weighting each estimate inversely to its variance. RVE, however, limits the weight given to an estimate by averaging the variances of all effect size estimates in a study, then using the inverse of this average as the weight (Hedges et al., 2010). Second, RVE uses a weighting scheme that models between-study heterogeneity in addition to within-study measurement error, which means smaller studies are not as heavily discounted, unlike UWLS. Hence, UWLS may serve as ‘conservative lower bound’ of summary effect sizes (Kaiser & Menkhoff, 2020). Third, misspecified weights in UWLS can lead to underestimation of standard errors and inflated Type-I error rates (López-López et al., 2017). RVE, in contrast, allows to estimate standard errors without making assumptions about the dependence structure as it proxies the study-specific variance-covariance structure of the errors by using products of the regression residuals. Though this can produce crude estimates, RVE standard errors are asymptotically valid (Pustejovsky & Tipton, 2022).

3.3.2 Ordinary Least Squares (OLS)

Using OLS, we address the dependence between effect sizes by clustering the standard errors at the cohort-by-program level. It is important to note that the OLS model assigns equal weights on each effect size estimate. This provides a general overview of the impact of VET on labor market outcomes, but it may not accurately reflect the actual "true effect" of VET on labor market outcomes (Kaiser & Menkhoff,

[15] Sensitivity analysis using different values of ρ lead to qualitatively similar results.

2020). In contrast, both UWLS and RVE are estimated by WLS where each of these two methods has its own way of weighting each effect size estimate, according to its precision (UWLS) or a combination of precision at the study level and between-study heterogeneity (RVE).

4. Results

Applying the inclusion criteria specified in section 2.1 led to a final sample of 70 studies, 77 clusters and 575 estimates belonging to three different strands of the literature (Table 1).

Table 1: Types of studies evaluating the labor market impact of VET

Study type	Focus and type of comparison group	Estimates	Cluster	Studies
1	Impact of VET versus general education (or vice versa)	317	40	39
2	Impact of VET compared to an unrestricted control group, i.e., that is not restricted in what it does	243	33	27
3	Impact of programs combining workplace- and classroom-based VET versus fully classroom-based VET programs	15	4	4
	Total	**575**	**77**	**70**

The empirical evidence of included studies belonging to the first strand of the literature ("Type 1" in Table 1) indicate that VET has a short- to medium-term advantage over general education. This includes studies employing experimental evaluation designs (Kemple & Willner, 2008), instrumental variables (IV) (Dougherty et al., 2019) or regression discontinuity designs (RDDs) (Brunner et al., 2023). In contrast, studies examining the long-term impacts suggest neutral to positive labor market impacts of VET versus general education. This includes papers employing difference-in-differences methods (DiD) (Oosterbeek & Webbink, 2007; Hall, 2016), IVs (Ferreyra et al., 2022) and RDDs (Malamud, 2010; Zilic, 2018; Silliman & Virtanen, 2022). Exceptions are Bertrand et al. (2021) using an RDD and Dai and Martins (2020), employing a shift-share instrument, who both find a positive long-term impact.

Studies belonging to the second strand of the literature evaluate the impact of VET versus an unrestricted control group ("Type 2" in Table 1), which may pursue any kind of outside activity (e.g., enter labor market, further education). Nearly two-thirds (63%) of these focus on programs at Community Colleges (CCs) in the US. We do not include estimates quantifying the effect of Certificate courses offered by CCs, because they are not comparable to Associate Degrees as these are often substantially shorter, have a different target group and are very heterogeneous in terms of duration and other characteristics between CCs.[16] Most of these studies employ person-level fixed-effects approach to isolate unobserved differences between degree earners and non-completers. Some studies add individual time trends to account for individual-specific heterogeneity that may change at a constant rate over time (Dynarski et al., 2018). Most studies find positive labor market effects of Associate Degrees (see e.g. Dynarski et al., (2018) for an overview). Studies not focusing on the US employing matching (Doerr, 2022), IV (Matthewes & Ventura, 2022), DiD estimation (Polidano & Ryan, 2016), RDD (Meneses et al., 2020) or experimental designs (Field et al., 2019) also find a positive effect.

Fewer studies belong to the third strand of the literature evaluating the impact of programs combining workplace- and classroom -based VET versus fully classroom -based VET programs ("Type 3" in Table

[16]We include estimates from one study (Carruthers C. K. & Sanford, 2018) that evaluates the impact of a two-year diploma in Tennessee, which is not an Associate Degree but comparable in duration and content breath.

1). Employing matching (Polidano & Tabasso, 2014) and IV estimation (Cavaglia et al., 2020; Bentolila et al., 2018; Parey, 2016), suggest a relative advantage of programs combining workplace- and classroom-based VET (Bolli et al., 2021). Given the low number of observations in type 3 studies, we restrict the empirical analysis of our paper to the first two types of studies.

Figure 1: Distribution of studies by publication year and study type

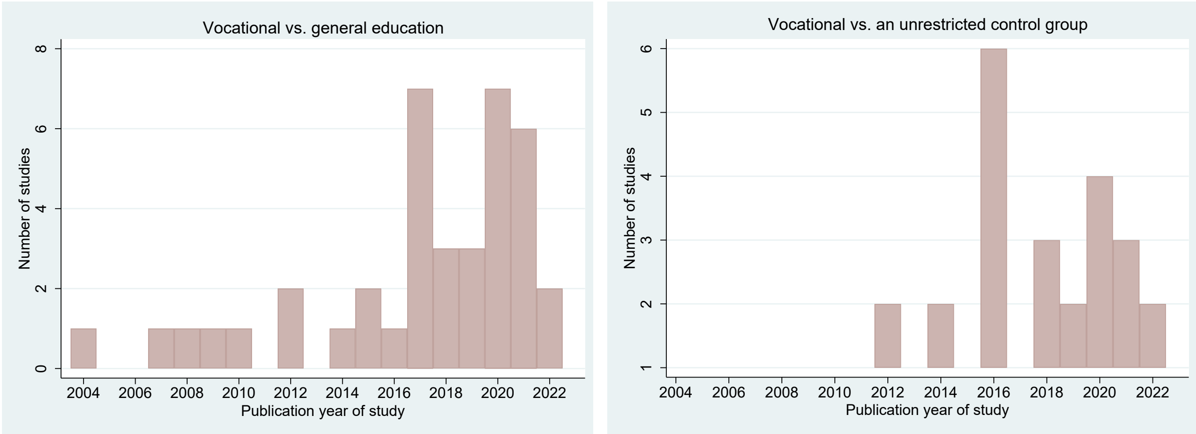


Figure notes: Both figures show the distribution of studies by publication year for type 1 studies (left) and type 2 studies (right).

Figure 1 depicts the distribution of studies by publication year for type 1 and 2 studies. For both types of studies, most were published after 2010 (see Table 2). The earliest of the type 1 and 2 studies was published in 2004 and 2012 respectively, and most studies were published in 2017 and 2020 for type 1 and in 2016 for type 2.

Table 2 shows descriptive statistics for studies included in our sample for both study types. Descriptive statistics at the cluster and program level are presented in Appendix B. More than half (51.3%) of the type 1 studies were conducted in Europe, while most (59.3%) type 2 studies focus on the US. Most type 1 (66.7%) and 2 (59.3%) studies were published in peer-reviewed publications and employ either an experimental design or a design without unconfoundedness assumption (type 1: 74.4%, type 2: 88.9%). While almost all studies of both types (type 1: 92.3%, type 2: 96.3%) evaluate the impact of vocational education on earnings-related outcomes, only about half do so for employment-related outcomes (type 1: 59%, type 2: 48.1%). Amongst these studies, most type 2 papers (77.8%) evaluate the impact on wages, and type 1 studies either wages or earnings (each 51.3%).[17] About half of the studies of both types show the short-run (0-5 years after program exit), fewer the medium-run (6-10 years) and even fewer type 2 studies the long-run (>10 years) impacts of VET. Most programs evaluated by type 2 studies (74.1%) do not last more than 24 months, while majority of programs in type 1 studies (48.7%) last between 25-36 months. This is due to the majority (67%) of type 2 studies evaluating the same US associate degree program. Therefore, duration of vocational education within the program also centers around 12-24 months for type 2 studies (81.5%), while it is more evenly distributed across categories for type 1 studies. The median age of vocational education students in both types of studies is split equally below 30 years and over 30 years.

[17]The difference between earnings and wages is that wages are conditional on being employed.

Table 2: Descriptive statistics

	Type 1 studies: VET vs. general education		Type 2 studies: VET vs. unrestricted control group	
Publication period:	**Studies**	**in %**	**Studies**	**in %**
2004-2010	5	12.8	2	7.4
2011-2019	19	48.7	16	59.3
2020-2022	15	38.5	9	33.3
Continent*:				
Africa	1	2.6	n.a.	n.a.
Asia	4	10.3	1	3.7
Europe	20	51.3	5	18.5
North America	8	20.5	16	59.3
Oceania	1	2.6	3	11.1
South America	4	10.3	2	7.4
Publication status:				
Peer-reviewed publication	26	66.7	16	59.3
Working paper	9	23.1	11	40.7
Technical report	2	5.1	n.a.	n.a.
Other publication type	2	5.1	n.a.	n.a.
Evaluation design:				
Experiment & Without unconfound.	29	74.4	24	88.9
With unconfound.	10	25.6	3	11.1
Outcome category:**				
Employment	23	59	13	48.1
Earnings	36	92.3	26	96.3
Outcome construct:**				
Employment probability/Participation rate	18	46.2	13	48.1
Hours worked	7	17.9	n.a.	n.a.
Earnings	20	51.3	6	22.2
Wage	20	51.3	21	77.8
Timing of follow-up measurement:**				
Program exit: 0-5 years ago	21	53.8	13	48.1
Program exit: 6-10 years ago	11	28.2	12	44.4
Program exit: >10 years ago	20	51.3	5	18.5
Program duration*:**				
<=24 months	11	28.2	20	74.1
>24 & <=36 months	19	48.7	5	18.5
>36 months	9	23.1	2	7.4
Duration vocational education*:**				
<=12 months	11	28.2	4	14.8
>12&<=24 months	14	35.9	22	81.5
>24 months	12	30.8	1	3.7
Participants:**				
Younger (<30 years)	22	56.4	11	40.7
Older (>=30 years)	24	61.5	16	59.3
Female	8	20.5	14	51.9
Male	15	38.5	13	48.1
Female and male together	30	76.9	15	55.6
Private sector involvement:				
Not involved	16	41	2	7.4
Involved	23	59	25	92.6
Program post-sec./tertiary:				
No	34	89.7	9	33.3
Yes	5	12.8	18	66.7
Learning place*****				
Classroom only	23	59	24	88.8
Combining workplace- and classroom	18	46	3	11.1

*Table notes: *Missing values for one study (Hanushek et al., 2017); **More observations due to overlapping categories at the cluster level; ***Observation missing due to missing information; ****The same program is evaluated by more than one study.*

Most studies of both types evaluate the impact of VET for men and women together rather than for both separately (type 1: 76.9%, type 2: 55.6%). Most programs involve private sector actors in design, implementation, or financing (type 1: 59%, type 2: 92.6%). While most type 1 studies (89.7%) analyze programs at the secondary education level, most type 2 studies (66.7%) do so for programs at the post-secondary or tertiary level. Lastly, almost half many (46%) type 1 studies and only few (11%) type 2 studies evaluate programs combining workplace- and classroom-based education.[18]

4.1 Assessing and accounting for publication bias

Publication bias occurs when choices about which studies to publish in a peer-reviewed journal are influenced by specific criteria such as the direction or statistical significance of results. This can distort the literature on the impact of VET on labor market outcomes by underrepresenting studies that do not meet the publication criteria based on which selection of studies takes place (Stanley & Doucouliagos, 2012). Accordingly, we test for the presence of this potential source of endogeneity.

4.1.1 The impact of VET vs. general education (Type 1)

Appendix C displays a funnel plot that illustrates the relationship between effect sizes and their precision, which is calculated as the inverse of their standard error. Without publication bias, the standard error of an estimate should be independent of the reported effect sizes, resulting in an inverted funnel shape plot that is symmetric around the true effect size (Stanley & Doucouliagos, 2012).[19]

As can be seen, most effect size estimates depicted in Appendix C are centered at the bottom of the funnel plot, depicting a slight skew of effect size estimates towards the lower left, and hence some presence of publication bias.

Table 3: FAT-PET model for type 1 studies

	All	Employment	Earnings	Experimental & without unconfound.	With unconfound.
Publ. Bias	1.190**	0.280	1.124*	0.651**	0.867*
	(0.568)	(0.423)	(0.656)	(0.313)	(0.470)
	[0.076, 2.304]	[-0.548, 1.108]	[-0.162, 2.410]	[0.037, 1.265]	[-0.054, 1.787]
Average effect	0.006	0.025*	0.001	0.007	0.074
	(0.007)	(0.015)	(0.007)	(0.005)	(0.081)
	[-0.008, 0.020]	[-0.004, 0.055]	[-0.013, 0.014]	[-0.004, 0.017]	[-0.085, 0.234]
Estimates	317	98	219	270	47
Cluster	40	25	35	29	11
Studies	39	23	36	29	10

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

Table 3 displays the outcomes from a regression-based test for publication bias using the SMD as dependent variable, as outlined by Stanley and Doucouliagos (2012) and detailed in Appendix C. The so-called FAT-PET model conducting a Funnel Asymmetry Test (FAT)" and a "Precision Effect Test (PET)" indicates the presence of publication bias in the sample including all outcomes (γ_{FAT} = 1.190**) and no statistically significant genuine effect beyond this bias (α_{PET}= 0.006). The same holds for sample splits for earnings-related outcomes and by identification strategy. Only the sample of employment-

[18] However, many CCs offer practical vocational education and training in a classroom setting.
[19] This implies that in the absence of publication bias, an inverted funnel shape is anticipated in the plot, where estimates with larger standard errors should appear at the bottom and be more spread out, while more precise estimates cluster at the top. Deviations to the left or right, or a skewed distribution of estimates, could also indicate the presence of publication bias (Stanley & Doucouliagos, 2012).

related outcomes depicts statistically significant evidence for a genuine effect, while publication bias does not seem to be present.

Given that we find broad evidence for publication bias and a genuine effect for more than half of clusters, we move forward with estimating the "Precision-Effect-Estimate with Standard Error (PEESE)" model (see Appendix C), which adjusts for publication bias more effectively than the FAT-PET model (Stanley & Doucouliagos, 2012). This adjustment slightly reduces the average estimated effect size in the univariate model for all outcomes from g = 0.023 in Table C1 in Appendix C to g = 0.021 in Table 5, both first column. Additionally, in Appendix D.1.2 and D.1.3, we conduct the FAT-PET test using WLS and OLS. Both show more evidence in favor of publication bias in the sample of type 1 studies. Therefore, we use the PEESE model in the further analysis of type 1 studies as suggested by Stanley and Doucouliagos (2012).

4.1.2 The impact of VET vs. an unrestricted control group (Type 2)

The funnel plot for type 2 studies, as shown in Figure C2 in Appendix C, shows a more dispersed distribution of estimates than the funnel plot for type 1 studies, with lower precision effect sizes in the bottom and fewer high precision estimates in the upper part of the distribution. Though the overall unweighted mean effect size (g=0.075, red vertical line) is positive, most high precision estimates are centered around zero.

The results of the FAT-PET model for type 2 studies depicted in Table 4 show evidence for publication bias (γ_{FAT} = 1.753***) and a statistically significant genuine effect beyond this bias (α_{PET}= 0.041***). The same holds for sample splits for employment- and earnings-related outcomes and by identification strategy.

Table 4: FAT-PET model for type 2 studies

	All	Employment	Earnings	Experimental & without unconfound.
Publ. Bias	1.753***	1.194***	2.109***	2.196***
	(0.415)	(0.356)	(0.779)	(0.353)
	[0.939, 2.566]	[0.497, 1.892]	[0.582, 3.636]	[1.505, 2.887]
Average effect	0.041***	0.048***	0.034**	0.037***
	(0.013)	(0.017)	(0.015)	(0.012)
	[0.017, 0.066]	[0.015, 0.081]	[0.006, 0.063]	[0.014, 0.060]
Estimates	243	62	181	214
Cluster	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

Therefore, we proceed by estimating PEESE models, which reduces the average estimated effect size in the univariate model for all outcomes from g = 0.075 in Table C2 in Appendix C to g = 0.071 in Table 7, both first column. Additionally, in Appendix D.2.2 and D.2.3, we conduct the FAT-PET test using WLS and OLS confirming the evidence in favor of publication bias in the sample of type 2 studies. Therefore, we use the PEESE model in the further analysis of type 2 studies as suggested by Stanley and Doucouliagos (2012).

4.2 Univariate Results

4.2.1 The impact of VET vs. general education

Figure 2 shows the distribution of the SMD Hedges' g for all outcomes along with the unweighted mean form an OLS model, and the PEESE RVE model. Both means (unrestricted=0.038; PEESE RVE=0.02) depict a relatively low impact of VET versus general education.

Figure 2: Distribution of the impact of VET versus general education, all outcomes

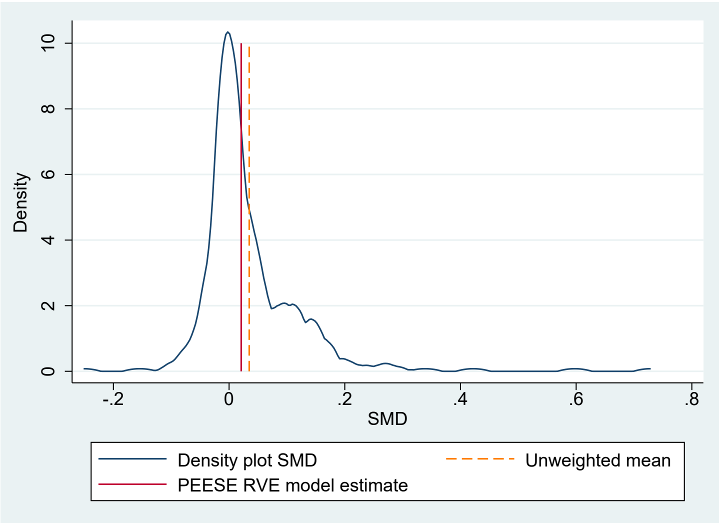


Figure notes: Figure 2 shows the distribution of the SMD for all outcomes along with the unweighted mean form an OLS model (dashed line, SMD=0.035, SD=0.089, Table D8 Appendix D.1.3) and the PEESE RVE model estimate (solid line, estimate=0.021, SE=0.007, see Table 5).

Table 5 shows univariate estimates of a PEESE RVE model of the impact of VET versus general education on labor market outcomes using the SMD as dependent variable, which controls for publication bias.

Table 5: Univariate Meta-Regression with Publication Bias (PEESE)

	All	Employment	Earnings	Experimental & without unconfound.	With unconfound.
Publ. Bias	4.257	2.044	4.345	4.601	2.086
	(2.615)	(4.701)	(2.826)	(5.568)	(1.339)
	[-0.869, 9.382]	[-7.170, 11.259]	[-1.193, 9.883]	[-6.313, 15.514]	[-0.539, 4.711]
Average effect	0.021***	0.030**	0.012**	0.012**	0.115
	(0.007)	(0.012)	(0.006)	(0.006)	(0.081)
	[0.007, 0.034]	[0.006, 0.054]	[0.001, 0.023]	[0.001, 0.023]	[-0.044, 0.274]
Estimates	317	98	219	270	47
Cluster	40	25	35	29	11
Studies	39	23	36	29	10

**p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD VAR.*

Results for the overall impact and most sample-splits are statistically significant, but considered as "small" compared to benchmarks provided by Kraft (2020): an average effect size of g=0.021 for all outcomes and a larger effect for employment- than for earnings-related outcomes.[20] Though not statistically significant, the impact of studies employing an evaluation design based on the

[20] As mentioned in section 3.1.1., according to Cohen (2013), SMDs ≤ 0.2 are considered small, while according to Kraft (2020) and Evans and Yuan (2020) SMDs ≤ 0.05 are considered as small.

unconfoundedness assumption, in our case mostly matching methods, depict an almost ten times higher average effect size than studies employing an experimental design or design without unconfoundedness assumption.

To get a first idea of how heterogeneity in program characteristics could impact results, Subfigures 3a, and 3b show the impact of VET versus general education on employment- and earnings-related outcomes respectively over time after an individual exited a program ("post-graduation"). As suggested by the literature, both subfigures show that the impact of VET versus general education is first positive and then tends towards zero over time.

Figure 3: Effect of vocational versus general education years post-graduation

a) Employment-related outcomes b) Earnings-related outcomes

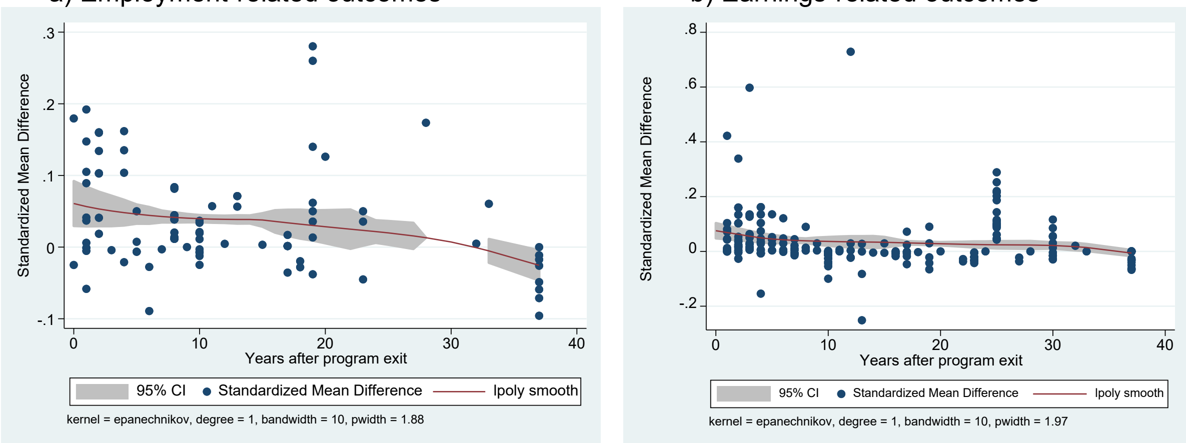


Figure notes: Both figures display the effect of vocational versus general education by years after an individual exited a program. Points represent estimates, the fitted line a kernel-weighted local polynomial regression ("lpoly smooth") along with the 95% confidence interval ("95% CI", shaded area).

Figure 4: Effect of vocational versus general education by program duration in months

a) Employment-related outcomes b) Earnings-related outcomes

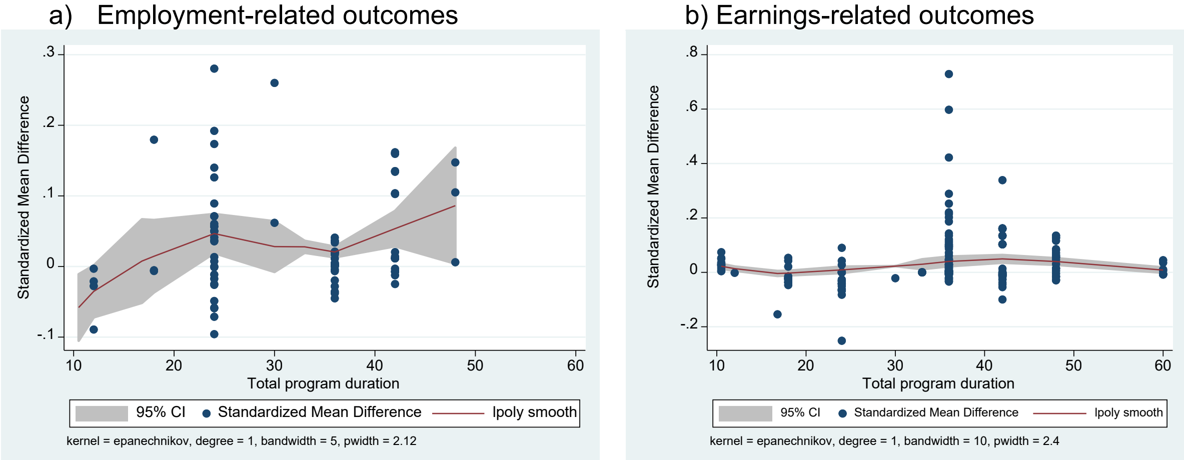


Figure notes: Both figures display the effect of vocational versus general education by program duration in months. Points represent estimates, the fitted line a kernel-weighted local polynomial regression ("lpoly smooth") along with the 95% confidence interval ("95% CI", shaded area).

Subfigures 4a/4b show the impact of VET versus general education on employment- and earnings-related outcomes by program duration in months.[21] As suggested by the human capital theory (Becker,

[21] Note that in most cases, the duration of the VET program and the general education program that are evaluated against one another is the same.

1962), one could expect that longer programs lead to relatively higher productivity and thus larger effect sizes. While subfigure 4a suggests that longer programs may lead to relatively larger employment-related effect sizes, subfigure 4b shows that effects on earnings are centered around zero over time.

Subfigures 5a/5b display the impact of VET versus general education on employment- and earnings-related outcomes by duration of vocational education within a given program in months, with the flipside being the within-program duration of general education, a measure for the within program intensity of VET. Both subfigures do not show evidence for a correlation between effect size magnitude and program intensity of VET.

Figure 5: Effect of vocational versus general education by duration of vocational education within the program (months)

a) Employment-related outcomes b) Earnings-related outcomes

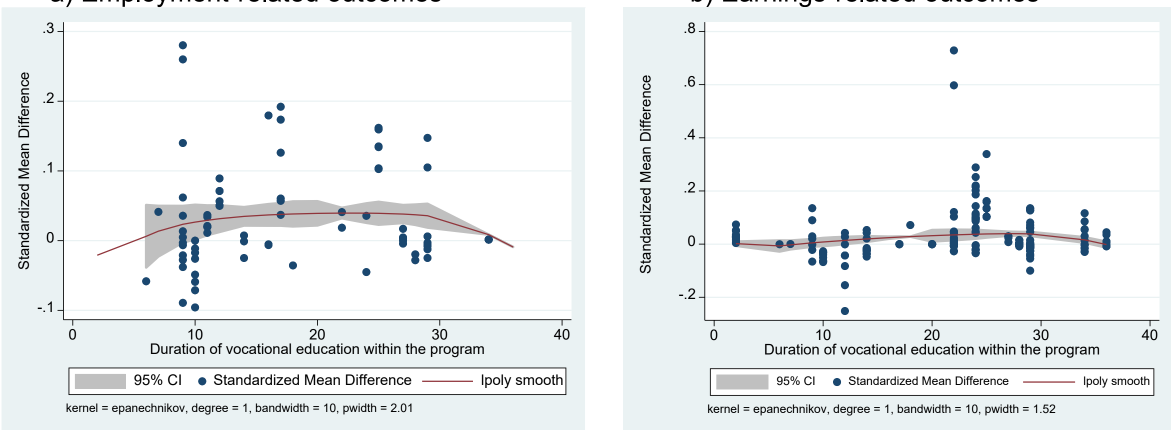


Figure notes: Both figures display the effect of vocational versus general education by duration of vocational education within the program in months. Points represent estimates, the fitted line a kernel-weighted local polynomial regression ("lpoly smooth") along with the 95% confidence interval ("95% CI", shaded area).

Employment probability and the logarithm of earnings/wages as outcome measures

Table 6 provides estimates of the "raw" effect of VET versus general education on the probability of being in employment and the logarithm of earnings/wages. Relative to general education, VET increases employment by about 4.8 percentage points. The relative impact on earnings/wages is not statistically significant.

Table 6: Univariate Meta-Regression with Publication Bias (PEESE)

	Employment probability	Logarithm of earnings/wages
Publ. Bias	-4.327	-1.533
	(4.915)	(5.874)
	[-13.959, 5.306]	[-13.046, 9.980]
Average effect	0.048***	0.108
	(0.019)	(0.078)
	[0.012, 0.085]	[-0.045, 0.261]
Estimates	63	160
Cluster	20	30
Studies	17	29

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets = Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

4.2.2 The impact of VET compared to an unrestricted control group

Figure 9 shows the distribution, the unweighted mean using OLS and the mean from a PEESE RVE model of the impact of VET compared to an unrestricted control group in terms of the SMD Hedges' g. Both means (unrestricted=0.122; PEESE RVE=0.085) are much larger than those of type 1 studies.

Figure 6: Distribution of the impact of VET versus un unrestricted control group

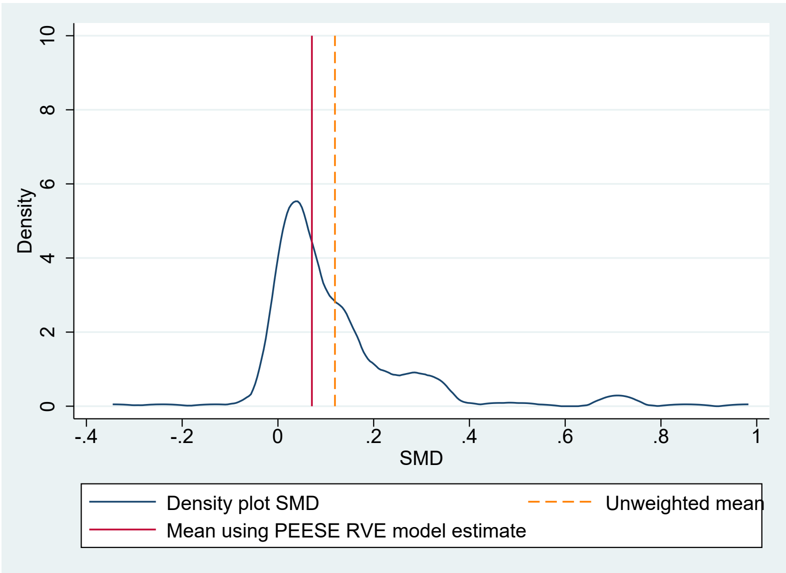


Figure notes: Figure 9 shows the distribution of the SMD for all outcomes along with the unweighted mean using OLS (dashed line, SMD=0.119, SD=0.161, see Table D20 in Appendix D.2.3) and the PEESE RVE model estimate (solid line, estimate=0.071, SE=0.013, see Table 7).

Table 7 shows a statistically significant impact of VET compared to an unrestricted control group for all outcomes and sample splits. The average impact for all outcomes amounts to a Hedges' g=0.071, with similar sizes observed for employment- (g=0.069) and earnings-related (g=0.068) outcomes. When excluding the three studies employing an evaluation design based on the unconfoundedness assumption, studies using experimental designs and designs without unconfoundedness assumption, the average effect depicts a Hedges' g of g=0.073. Overall, the impact is larger for type 2 than type 1 studies. According to standards provided by Kraft (2020), these effect sizes can be considered as "medium".

Table 7: Univariate Meta-Regression with Publication Bias (PEESE)

	All	Employment	Earnings	Experimental & without unconfound.
Publ. Bias	4.236***	3.980***	3.955	5.206***
	(1.027)	(0.738)	(2.958)	(0.775)
	[2.224, 6.248]	[2.534, 5.426]	[-1.843, 9.753]	[3.688, 6.725]
Average effect	0.071***	0.069***	0.068***	0.073***
	(0.013)	(0.017)	(0.013)	(0.014)
	[0.046, 0.096]	[0.037, 0.102]	[0.042, 0.095]	[0.046, 0.100]
Estimates	243	62	181	214
Cluster	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD. There are only three studies and clusters (27 estimates) employing an evaluation design based on the unconfoundedness assumption (matching methods).*

As in the previous section, we show the impact of VET versus an unrestricted control group on employment- and earnings-related outcomes respectively over time after an individual exited a program we show in Subfigure 7a, and 7b. Both subfigures show that the impact of VET versus an unrestricted control group is first positive and then tends towards zero over time.

Figure 7: Effect of vocational education by years post-graduation

a) Employment-related outcomes b) Earnings-related outcomes

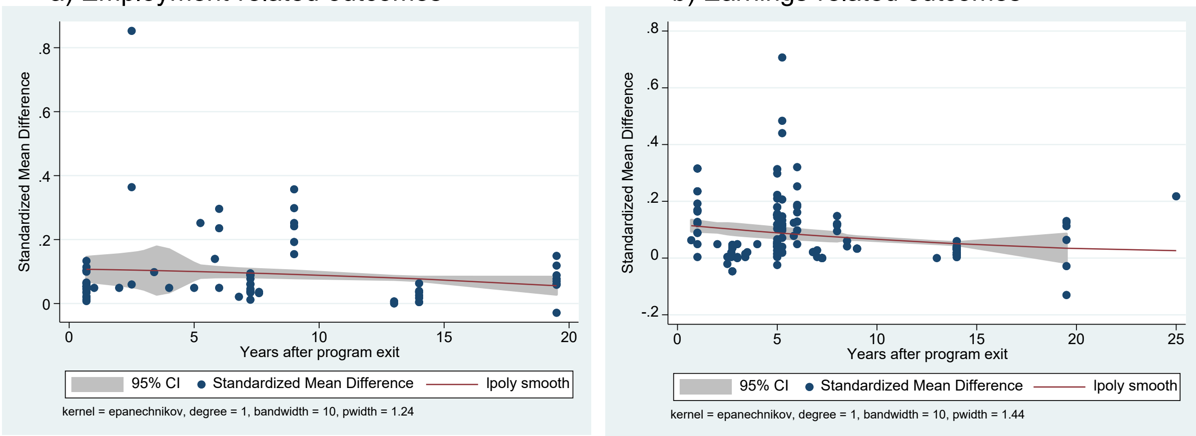


Figure notes: Both figures display the effect of vocational education by years after an individual exited a program. Points represent estimates, the fitted line a kernel-weighted local polynomial regression ("lpoly smooth") along with the 95% confidence interval ("95% CI", shaded area).

Subfigures 8a and 8b show the impact of VET versus an unrestricted control group on employment- and earnings-related outcomes by program duration in months. According to both subfigures, longer programs do not lead to higher effect sizes.

Figure 8: Effect of vocational versus an unrestricted control group by program duration in months

a) Employment-related outcomes b) Earnings-related outcomes

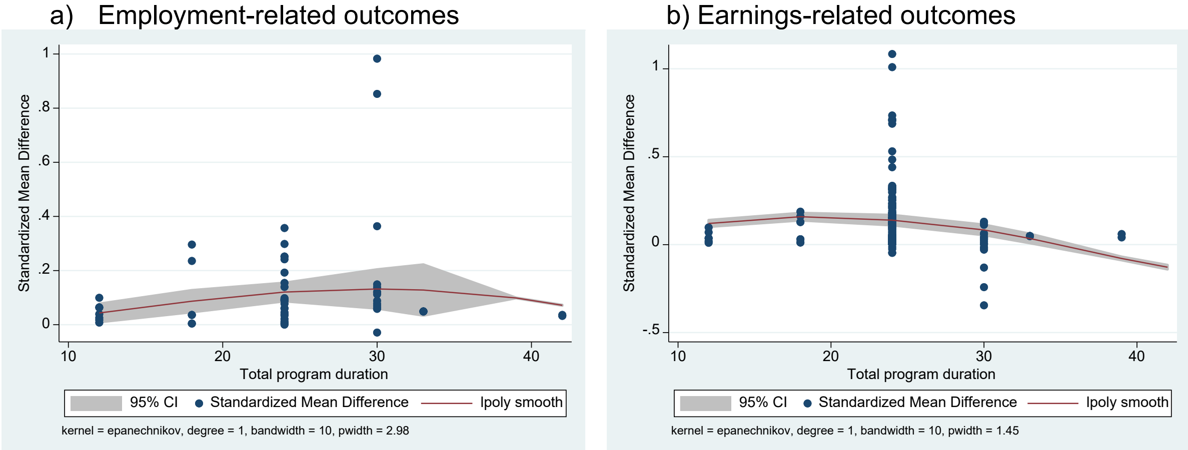


Figure notes: Both figures display the effect of vocational versus an unrestricted control group by program duration in months. Points represent estimates, the fitted line a kernel-weighted local polynomial regression ("lpoly smooth") along with the 95% confidence interval ("95% CI", shaded area).

Subfigures 9a and 9b show the impact of VET versus an unrestricted control group on employment- and earnings-related outcomes by duration of vocational education within a given program in months, i.e. the intensity of VET.

Both subfigures show that effect sizes decrease by duration of vocational education.

Figure 9: Effect of vocational education by duration of vocational education within the program (months)

a) Employment-related outcomes b) Earnings-related outcomes

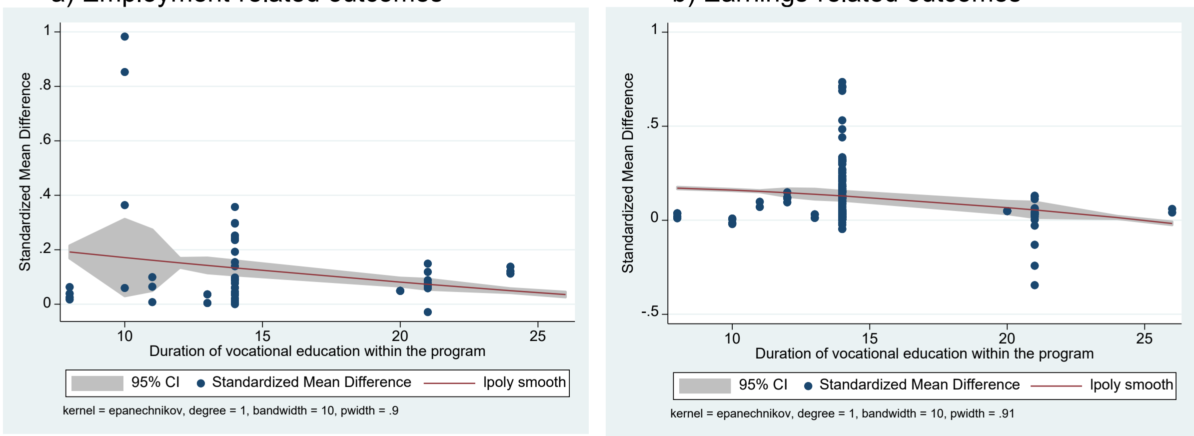


Figure notes: Both figures display the effect of vocational education by duration of vocational education within the program in months. Points represent estimates, the fitted line a kernel-weighted local polynomial regression ("lpoly smooth") along with the 95% confidence interval ("95% CI", shaded area).

Employment probability and the logarithm of earnings/wages as outcome measures

Table 8provides estimates of the "raw" effect of VET compared to an unrestricted control group on the probability of being employed and the logarithm of earnings/wages. Relative to the control group, VET increases employment by about 11.7 percentage points and earnings/wages by 37.6 percentage points.

Table 8: Univariate Meta-Regression with Publication Bias (PEESE)

	Employment probability	Logarithm of earnings/wages
Publ. Bias	0.414	-10.214
	(1.038)	(8.103)
	[-1.620, 2.448]	[-26.096, 5.669]
Average effect	0.117***	0.376***
	(0.014)	(0.097)
	[0.090, 0.144]	[0.185, 0.566]
Estimates	66	122
Clusters	19	24
Studies	13	20

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets = Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

The impact of vocational education at Community Colleges (CCs)

Of the studies evaluating the impact of VET versus and unrestricted control group, nearly two-thirds (63%) evaluate the impact of vocational education at Community Colleges (CCs) in the U.S. To check if this group is driving the main results, we now only include estimates of the effect of Associate Degrees at CCs (see section 4.0). Therefore, we analyze how the average impact of VET for the sample of studies evaluating Associate Degrees at CCs differs from the effect using the remaining studies. The average impact of VET at CCs (g=0.054, Table D25 in Appendix D.2.4) is significantly higher than for the remaining studies (g=0.039).

Since 11 of the 17 studies on CCs analyze how the impact of VET at CCs differs by field of study, we also show average impacts by field of study, which may be particularly interesting for policy makers. Table 9 shows the resulting mean and standard deviation by field of study using the SMD Hedges' g as an outcome for all 11 studies in the left column and using the logarithm of earnings/wages as an outcome based on eight studies in the right column. Though the results in Table 9 are based on a rather low number of studies, they provide indicative evidence that outcomes are highest for the study fields Health Science (g=0.207, log. earn/wages=0.547) and Agriculture (g=0.16, log. earn/wages=0.247), suggesting wage returns of about 54.7% and 24.7% respectively.[22]

Table 9: Descriptive statistics - Variation in mean outcomes by field of study

		SMD Hedges' g			Logarithm of earnings/wages		
Field number	**Occupational field**	**Obs**	**Mean**	**Std. dev.**	**Obs**	**Mean**	**Std. dev.**
1	**Human Services & Education** - Government & Public Administration - Human Services - Education & Training - Law, Public Safety, Corrections, & Security	28	0.080	0.113	25	0.274	0.767
2	**Communication & Information Systems** - Communication Arts - Information Technology	17	0.074	0.100	14	0.129	0.066
3	**Health Sciences**	34	0.207	0.247	24	0.547	0.839
4	**Skilled Technical Sciences** - Manufacturing - Energy & Engineering - Transportation, Distribution, & Logistics - Architecture & Construction	38	0.145	0.211	22	0.268	1.045
5	**Agriculture**	8	0.16	0.123	6	0.247	0.101
6	**Business, Marketing, & Management** - Manufacturing - Energy & Engineering - Transportation, Distribution, & Logistics - Architecture & Construction	32	0.097	0.199	20	0.065	0.222
Estimates			157			111	
Cluster			11			8	
Studies			11			8	

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

4.3 Robustness checks

4.3.1 Multivariate meta-regression

A limitation of univariate regressions is that they may fail to account for correlations between key study or program characteristics. To address this issue, we estimate multivariate meta-regressions that allow us to control jointly for a broad array of observable characteristics. As suggested by Stanley and Doucouliagos (2012), for instance, we include different sets of independent variables in a cascading manner, while the last column includes all covariates.

[22] As shown in Table D26 in Appendix D.2.4, a RVE PEESE regression using the field of study Human Services & Education as baseline confirms that outcomes for the fields Health Science and Agriculture are statistically significantly different from the baseline.

The impact of VET vs. general education

Table 10 shows results of a multivariate regression using all outcomes for the impact of VET versus general education, Hedges' g as an outcome variable, and accounts for publication bias. Tables D1 and D2 in Appendix D display results of multivariate regressions using only employment- or earnings-related outcomes.

The only statistically significant finding is that studies employing experimental designs or designs without unconfoundedness assumption depict lower average effect sizes than studies using designs based on the unconfoundedness assumption (in our case matching methods). Though not statistically significant, the sign of coefficients measuring post-program outcomes after 6-10 and more than 10 years relative to 0-5 years after program exit support the graphical findings in section 4.2.1, Figure 3, suggesting that the impact of VET tends to be lower in the long run.

Table 10: RVE PEESE regression using all outcomes and SMD

	1	2	3	4	5	6
Var (SMD)	2.930	2.559*	2.677*	2.635	2.490*	2.152*
	(1.788)	(1.512)	(1.608)	(1.657)	(1.462)	(1.301)
Exper. & Without unconfound.	-0.078**	-0.083*	-0.083*	-0.094**	-0.082*	-0.101*
Baseline: with unconfound.	(0.034)	(0.047)	(0.043)	(0.039)	(0.046)	(0.056)
Female participants		-0.010			-0.013	-0.010
Baseline: Pooling fem. & male		(0.010)			(0.010)	(0.016)
Male participants		0.037			0.040	0.034
Baseline: Pooling fem. & male		(0.030)			(0.032)	(0.038)
Average age		-0.001*			-0.001	-0.001
		(0.001)			(0.001)	(0.001)
Program exit: 6-10 years ago			-0.002		0.001	-0.016
Baseline: up to 5 years			(0.014)		(0.011)	(0.019)
Program exit: >10 years ago			-0.006		0.008	0.014
Baseline: up to 5 years			(0.018)		(0.021)	(0.024)
Program durat.: 13-24 months				-0.017		-0.027
Baseline: up to 12 months				(0.034)		(0.039)
Program durat.: 25-36 months				0.014		0.020
Baseline: up to 12 months				(0.030)		(0.026)
Program at tertiary level				0.019		0.024
Baseline: at secondary level				(0.023)		(0.041)
Constant	0.091***	0.130**	0.101**	0.095**	0.133**	0.137**
	(0.033)	(0.054)	(0.041)	(0.048)	(0.057)	(0.066)
Obs	317	317	317	306	317	306
Cluster	40	40	40	37	40	37
Studies	39	39	39	38	39	38

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

The impact of VET vs. an unrestricted control group

Table 11 presents the outcomes of a multivariate regression that analyzes the impact of VET versus an unrestricted control group. Tables D13 and D14 in Appendix D display results of multivariate regressions using only employment- or earnings-related outcomes. As for type 1 studies, results for type 2 studies suggest that studies employing experimental designs or designs without unconfoundedness assumption depict lower average effect sizes than studies using designs based on the unconfoundedness assumption. Again, though not statistically significant, coefficients measuring post-program outcomes after 6-10 and more than 10 years relative to 0-5 years after program exit support the graphical findings in section 4.2.2, Figure 7, suggesting that the impact of VET tends to be lower in the long run.

Table 11: RVE PEESE regression using all outcomes and SMD

	1	2	3	4	5	6
Var (SMD)	4.308***	4.123***	4.219***	4.021***	4.083***	3.948***
	(0.948)	(0.907)	(0.880)	(0.893)	(0.853)	(0.892)
Exper. & Without unconfound.	0.033**	0.043**	0.039**	0.020	0.045**	-0.011
Baseline: with unconfound.	(0.016)	(0.021)	(0.016)	(0.029)	(0.020)	(0.045)
Female participants		0.010			0.010	0.004
Baseline: Pooling fem. & male		(0.029)			(0.029)	(0.033)
Male participants		-0.011			-0.007	0.001
Baseline: Pooling fem. & male		(0.031)			(0.035)	(0.040)
Average age		-0.001			-0.000	0.001
		(0.002)			(0.002)	(0.002)
Program exit: 6-10 years ago			0.003		0.005	-0.001
Baseline: up to 5 years			(0.029)		(0.034)	(0.039)
Program exit: >10 years ago			-0.029		-0.025	-0.017
Baseline: up to 5 years			(0.038)		(0.047)	(0.065)
Program durat.: *0-12 months*				-0.041		-0.029
Baseline: 25-36 months				(0.048)		(0.074)
Program durat.: 13-24 months				-0.001		0.029
Baseline: 25-36 months				(0.048)		(0.050)
Program at tertiary level				0.034		0.038
				(0.043)		(0.056)
Constant	0.040***	0.052	0.040**	0.037***	0.038	0.004
	(0.008)	(0.051)	(0.018)	(0.012)	(0.051)	(0.066)
Obs	243	235	243	237	235	229
Cluster	33	31	33	32	31	30
Studies	27	26	27	26	26	25

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

4.3.2 Re-estimating main results by UWLS and OLS

We test the robustness of our results by re-estimating all main results using UWLS and OLS. As highlighted by Kaiser & Menkhoff (2020), UWLS estimates may provide a ‘conservative lower bound’ of summary effect sizes. This is also the case in our paper, where estimations from univariate UWLS models are smaller than those of RVE or OLS. Results of univariate OLS models are closer to RVE estimates than UWLS. This also holds for results of multivariate meta-regressions. Overall, the results are qualitatively consistent across all three statistical approaches. This suggests that the choice of method—whether UWLS, OLS, or RVE—does not significantly influence the outcomes, providing confidence that the findings are robust and reliable regardless of the estimation technique employed.

5. Summary and conclusion

In this meta-analysis, we systematically review and summarize the empirical impact of *formal* VET on labor market outcomes. We include studies employing a rigorous causal counterfactual evaluation design published 1990-2022.

Through our extensive online search, we identified three different types of studies: 39 studies comparing the impact of VET to general education, 27 studies contrasting VET with an unrestricted control group, and four studies evaluating the effects of work-based versus school-based VET. Only the first two types of studies allow a meaningful empirical analysis.

To aggregate effect sizes from these studies, we employ the SMD Hedges' g (Hedges, 1981). To compute the economic impact of VET, we also show results quantifying its impact on employment probability and earnings/wages for a subset of studies. We employ Robust Variance Estimation (Hedges et al., 2010) to account for dependencies within and between studies.

Based on a sample of 39 studies, we find that the overall impact of VET versus general education (Type 1 studies) is positive but small (Hedges' g= 0.021***) and tends to zero over time, though more studies evaluate the short- to medium-term effects. About half of the studies in this sample can be used to compute the economic impact of VET versus general education in terms of employment. This is also small (4.8%***) and not statistically significant in terms of earnings/wages. These results are consistent with qualitative summaries of the literature based on studies with experimental or quasi-experimental evaluation designs, refuting the hypothesis of a relative advantage of general education in the long run (e.g., Silliman & Virtanen (2022), McNally et al. (2022)). Studies using more rigorous counterfactual evaluation designs show even more relatively lower average impacts. These results mostly apply to VET programs at the upper-secondary level, since 90% of studies evaluate programs at this level.

Our analysis also shows that, based on a sample of 27 studies, the impact of VET versus an unrestricted control group (Type 2 studies) is positive and to be considered as “medium” (Hedges' g=0.071***) when compared to field-specific benchmarks (Kraft, 2020). Using about half of all studies in this sample, we observe that VET increases employment by 11.7%*** and earnings/wages by 37.6%*** relative to an unrestricted control group. About two-thirds of studies evaluate Associate Degree programs at Community Colleges in the US. These depict higher average effect sizes than the remaining studies. Analyzing the impact of Associate Degree programs by field reveals that returns are highest for the fields Health Science and Agriculture, yielding 54.7%*** and 24.7%*** higher earnings/wages than the control group. In general, the number of studies analyzing other than CCs programs, especially at the upper-secondary level rather than the post-secondary, tertiary level is rather low.

In conclusion, this meta-analysis provides a comprehensive assessment of the effects of formal VET on labor market outcomes across different educational levels and geographic contexts. Our results are the first quantitative summary of the literature. It provides numerical evidence for both strands of the literature evaluating the impact of VET on labor market outcomes and thus goes beyond the qualitative summaries that have existed so far. We do not find evidence that other program or study characteristics lead to higher effects sizes, i.e., effect size heterogeneity remains limited.

To date, many studies have focused purely on classroom-based VET programs. The low effect sizes of VET- relative to general education or to an unrestricted control group- could be explained by the fact that the program quality of classroom-based VET programs can vary greatly depending on whether they are based on a uniform national qualification standard or were developed by each college based on school autonomy. Finally, the close coordination between actors in the labor market and education system also plays an important role for a smooth transition into the labor market. In the case of classroom-based VET programs, it must be assumed that there is little coordination with the skills in demand on the labor market. This in contrast to programs combining classroom- and workplace-based VET programs.[23]

A major limitation of our study is that the quality of our results is only as good as the included studies, even if we limit our evidence base to studies with the most rigorous evaluation designs possible. There is still a need to increase the number of high-quality type 1 and 2 studies, especially those providing empirical evidence on the long-term effects of VET.

[23] Multivariate meta-regressions contrasting results for programs combining workplace- and classroom-based education with those of purely classroom-based ones do not show a significant difference in effect sizes.

References

Balasubramanian, P., Ibanez, M., Khan, M., & Sahoo, S. (2024). Does women's economic empowerment promote human development in low-and middle-income countries? A meta-analysis. . *World Development*, 178, 106588.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy, 70*, 9–49.

Bentolila, S., Cabrales, A., & Jansen, M. (2018). El impacto de la FP dual sobre la inserción lavoral de los jóvenes: evidencia para la comunidad de Madrid. *FEDEA, No. eee2018-20*.

Bertrand, M., Mogstad, M., M., & Mountjoy, J. (2021). Improving educational pathways to social mobility: evidence from Norway's reform 94. *Journal of Labor Economics, 39*(4), 965-1010.

Bolli, T., Oswald-Egg, M., & Rageth, L. (2021). Meet the need–The role of vocational education and training for the youth labour market. *Kyklos*, 74(3), 321-348.

Borenstein, M. (2009). Effect sizes for continuous data. En H. Cooper, L. Hedges, & J. Valentine, *The handbook of research synthesis and meta-analysis* (págs. 221-235.). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods, 1*, 97–111.

Brunner, E., Dougherty, S., & Ross, S. (2023). The effects of career and technical education: Evidence from the Connecticut Technical High School System. *Review of Economics and Statistics, 105*(4), 867-882.

Carruthers C. K., &., & Sanford, T. (2018). Way station or launching pad? Unpacking the returns to adult technical education. *Journal of Public Economics*, 165, 146-159.

Carruthers, C., & Jepsen, C. (2021). Vocational Education: An international perspective. En *The Routledge Handbook of the Economics of Education* (págs. 343-380).

Cavaglia, C., McNally, S., & Ventura, G. (2020). Do apprenticeships pay? Evidence for England. *Oxford Bulletin of Economics and Statistics, 82*(5), 1094-1134.

Cheung, M. W.-L. (2019). A Guide to Conducting a Meta-Analysis with Non-Independent Effect Sizes. *Neuropsychology Review, 29*, 387–396. https://doi.org/10.1007/s11065-019-09415-6

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* New York: Academic press.

Dai, L., & Martins, P. (2020). Does vocational education pay off in China? Instrumental-variable quantile-regression evidence. *GLO Discussion Paper, No. 495*.

Doerr, A. (2022). Vocational training for female job returners-Effects on employment, earnings and job quality. *Labour Economics, 75*(102139).

Dougherty, S. M., & Ecton, W. (2021). The economic effect of vocational education on student outcomes. En *Oxford Research Encyclopedia of Economics and Finance.* https://doi.org/https://doi.org/10.1093/acrefore/9780190625979.013.656

Dougherty, S., Gottfried, M., & Sublett, C. (2019). Does increasing career and technical education coursework in high school boost educational attainment and labor market outcomes? *Journal of Education Finance, 44*(4), 423-447.

Dynarski, S., Jacob, B., B., & Kreisman, D. (2018). How important are fixed effects and time trends in estimating returns to schooling? Evidence from a replication of Jacobson, Lalonde, and Sullivan, 2005. *Journal of Applied Econometrics, 33*(7), 1098-1108.

Evans, D. K., & Yuan, F. (2020). How Big Are Effect Sizes in International Education Studies. *Center for Global Development, Working Paper, 545*.

Ferreyra, M. M., Galindo, C., C., & Urzúa, S. (2022). Labor Market Effects of Short-Cycle Higher Education Programs: Lessons from Colombia. *National Bureau of Economic Research (NBER), No. w30178*.

Field, E. M., Linden, L., Malamud, O., Rubenson, D., & Wang, S. (2019). Does vocational education work? Evidence from a randomized experiment in Mongolia. *National Bureau of Economic Research (NBER), No. w26092*.

Hall, C. (2016). Does more general education reduce the risk of future unemployment? Evidence from an expansion of vocational upper secondary education. *Economics of Education Review, 52*, 251-271.

Hampf, F., Wiederhold, S., & Woessmann, L. (2017). Skills, earnings, and employment: exploring causality in the estimation of returns to skills. *Large-scale Assessments in Education, 5*, 1–30.

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics, 6*, 107–128.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust Variance Estimation in Meta-Regression with Dependent Effect Size Estimates. *Research Synthesis Methods, 1*, 39–65.

Imbens. (2023). Causal Inference in the Social Sciences. *Annual Review of Statistics and Its Application*, 11.

Imbens, G., & Wooldridge, J. (2007). What's new in econometrics: estimation of average treatment effects under unconfoundedness. Lecture Notes, 1. *NBER, Summer 2007. Lecture Notes, 1.*

Kaiser, T., & Menkhoff, L. (2020). Financial Education in Schools: A Meta-Analysis of Experimental Studies. *Economics of Education Review*. https://doi.org/https://doi.org/10.1016/j.econedurev.2019.101930

Kaiser, T., Lusardi, A., Menkhoff, L., & Urban, C. (2022). Financial Education Affects Financial Knowledge and Downstream Behaviors. *Journal of Financial Economics*, 145(2), 255-272. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.09.022

Kemper, J., Stöterau, J., & Ghisletta, A. (2022). The Impact of Vocational Training Interventions on Youth Labor Market Outcomes: A Meta-Analysis. Obtenido de https://ssrn.com/abstract=4217580 or http://dx.doi.org/10.2139/ssrn.4217580

Kemple, J., & Willner, C. (2008). *Technical Resources for Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood. .* MDRC.

Kluve, J., Puerto, S., Robalino, D., Rom, J., Rother, F., Stöterau, J., . . . Witte, M. (2019). Do youth employment programs improve labor market outcomes? A quantitative review. *World Development, 114*, 237-253.

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher, 49*, 241–253. https://doi.org/10.3102/0013189X20912798

Kvarven, A., Strømland, E., & Johannesson, M. (April de 2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour, 4*, 423–434. https://doi.org/10.1038/s41562-019-0787-z

López-López, J. A., Van den Noortgate, W., Tanner-Smith, E., Wilson, S. J., S., & Lipsey, M. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A Monte Carlo simulation. *Research Synthesis Methods*, 8(4), 435-450.

Malamud, O. &.-E. (2010). General education versus vocational training: Evidence from an economy in transition. *The Review of Economics and Statistics, 92*(1), 43-60.

Matthewes, S. H., & Ventura, G. (2022). Labour Market Returns to Vocational Education in the Presence of Multiple Alternatives.

McNally, S., Ventura, G., & Virtanen, H. (2022). Returns to vocational education and training. En *Handbook of Labor, Human Resources and Population Economics* (págs. 1-30). Cham: Springer International Publishing.

Meneses, F., Blanco, C., C., Flores, R., & Paredes, R. (2020). Conflicting Incentives: Government Financial Aid, Vocational-to-University Track Change and Graduates' Wages in Chile. *CAF - WORKING PAPER, # 2020/08*.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings.* Cambridge, MA: NBER.

OECDa. (2023). *Education at a Glance 2023: OECD Indicators.* Paris: OECD Publishing. https://doi.org/https://doi.org/10.1787/e13bef63-en

OECDb. (2023). *Building Future-Ready Vocational Education and Training Systems, OECD Reviews of Vocational Education and Training.* Paris: OECD Publishing. https://doi.org/https://doi.org/10.1787/28551a79-en

Oosterbeek, H., & Webbink, D. (2007). Wage effects of an extra year of basic vocational education. *Economics of Education Review, 26*(4), 408-419.

Parey, M. (. (2016). Vocational Schooling versus Apprenticeship Training. Evidence from Vacancy Data. *Beiträge zur Jahrestagung des Vereins für Socialpolitik, No. F20-V1*.

Polidano, C., & Ryan, C. (2016). Long-term outcomes from Australian vocational education. *Melbourne Institute Working Paper , No. 35/16*.

Polidano, C., & Tabasso, D. (2014). Making it real: The benefits of workplace learning in upper-secondary vocational education and training courses. *Economics of Education Review, 42*, 130-146.

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23(3)*, 425-438.

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods, 26*, 141–160. https://doi.org/10.1037/met0000300

Ryan, P. (2012). Apprenticeship: between theory and practice, school and workplace. En M. Pilz, *The future of vocational education and training in a changing world* (págs. 402-432). Springer VS Verlag für Sozialwissenschaften Wiesbaden.

Silliman, M., & Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics, 14*(1), 197-224.

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

Stanley, T. D., & Doucouliagos, H. (2017). Neither Fixed Nor Random: Weighted Least Squares Meta-Regression. *Research synthesis methods, 8*, 19–42.

Stanley, T. D., Doucouliagos, H., & Ioannidis, J. P. (2022). Beyond Random Effects: When Small-Study Findings Are More Heterogeneous. *Advances in Methods and Practices in Psychological Science, 5*, 25152459221120427. https://doi.org/10.1177/25152459221120427

Tipton, E. (2013). Robust Variance Estimation in Meta-Regression with Binary Dependent Effects. *Research synthesis methods, 4*, 169–187.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. . *Psychological methods, 20(3)*, 375.

UNESCO Institute for Statistics. (2012). *International standard classification of education: ISCED 2011.* Montreal, Quebec: UNESCO. https://doi.org/http://dx.doi.org/10.15220/978-92-9189-123-8-en

Waheed, H. (2023). Nudging smokers away from lighting up: A meta-analysis of framing effect in current smokers. *Journal of Behavioral and Experimental Economics*, 104, 101998.

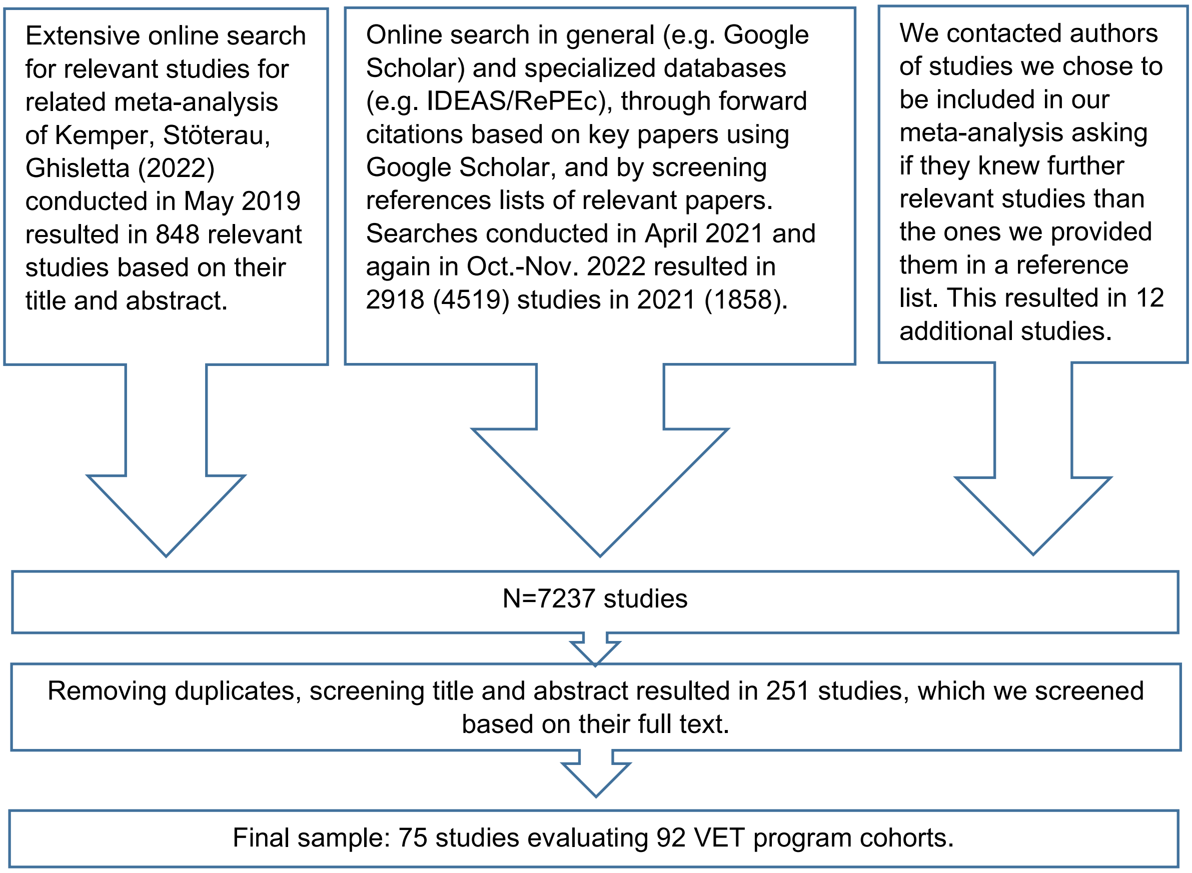
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

Zilic, I. (2018). General versus vocational education: Lessons from a quasi-experiment in Croatia. *Economics of Education Review, 62*, 1-11.

Appendix

A. Search for relevant studies

Figure A1: Strategy to sample relevant studies



B. Detailed descriptives

Table B1: Detailed descriptives

	VET vs. general education				VET vs. unrestricted control group			
Publication period:	**Estimates**	**Cluster**	**Programs**	**Studies**	**Estimates**	**Cluster**	**Programs**	**Studies**
2004-2010	41	4	4	5	18	2	2	2
2011-2019	123	19	15	19	149	21	7	16
2020-2022	153	17	12	15	76	10	7	9
Continent*:								
Africa	2	1	1	1	n.a.	n.a.	n.a.	n.a.
Asia	26	4	3	4	8	2	2	1
Europe	195	21	16	20	29	5	4	5
North America	48	6	2	8	153	20	2	16
Oceania	6	1	1	1	40	4	3	3
South America	29	4	4	4	13	2	2	2
Publication status:								
Peer-reviewed publication	240	29	20	26	162	20	6	16
Working paper	48	9	9	9	81	13	9	11
Technical report	24	1	1	2	n.a.	n.a.	n.a.	n.a.
Other publication type	5	1	1	2	n.a.	n.a.	n.a.	n.a.
Evaluation design:								
Experiment & Without unconfound.	270	29	21	29	214	30	12	24
With unconfound.	47	11	10	10	29	3	3	3
Outcome category:**				0				
Employment	98	25	20	23	62	19	11	13
Income	219	35	25	36	181	31	12	26
Outcome construct:**								
Employment probability/Participation rate	70	22	18	18	62	19	11	13
Hours worked	28	6	6	7	n.a.	n.a.	n.a.	n.a.
Earnings	119	20	13	20	32	10	4	6
Wage	100	21	18	20	149	23	9	21
Timing of follow-up measurement:**								
Program exit: 0-5 years ago	106	21	16	21	112	14	6	13
Program exit: 6-10 years ago	55	13	11	11	84	16	5	12
Program exit: >10 years ago	156	19	16	20	47	6	5	5
Program duration*:**								
<=24 months	90	12	11	11	185	25	7	20
>24 & <=36 months	117	18	13	19	46	5	4	5
>36 months	99	8	6	9	6	2	1	2
Duration vocational education*:**								
<=12 months	75	11	10	11	26	4	4	4
>12&<=24 months	93	15	11	14	205	26	7	22
>24 months	130	10	7	12	2	1	1	1
Participants:**								
Younger (<25 years)	123	21	17	22	111	14	3	11

Older(>=25 years)	194	26	19	24	132	19	12	16
Female	43	8	8	8	97	18	8	14
Male	77	15	13	15	67	14	8	13
Female and male together	197	31	23	30	79	17	9	15
Private sector involvement:								
Not involved	165	20	14	16	14	2	2	2
Involved	152	20	17	23	229	31	11	25
Program post-sec./tertiary:								
No	294	35	26	34	66	10	8	9
Yes	23	5	5	5	177	23	5	18

*Table notes: *Obs missing for one study (Hanushek et al., 2017); **More obs due to overlapping categories at the cluster level; ***Obs missing due to missing information.*

C. Testing for publication bias

We conduct formal tests for publication bias using regression methods as outlined by Stanley and Doucouliagos (2012). Initially, we incorporate the standard error SE_{ij} of the effect size estimates into the univariate RVE meta-regression model to perform the "Funnel Asymmetry Test (FAT)", which accounts for publication bias (H_0: γ_{FAT} = 0). Concurrently, we apply the "Precision Effect Test" (PET) within the same model to assess the existence of a true effect, separate from publication bias ($H_0 : \alpha_{PET} = 0$).

$$g_{ij} = \alpha_{PET} + SE_{ij}\,\gamma_{FAT} + v_j + \varepsilon_{ij} \qquad (3)$$

If the FAT-PET results suggest both a statistically significant publication bias and a genuine effect, one can proceed by employing the “Precision-Effect-Estimate with Standard Error (PEESE)” model to obtain unbiased estimates of the genuine effect. According to Stanley and Doucouliagos (2012), the PEESE estimator is more effective in approximating the underlying "true" effect when the FAT-PET tests are rejected. In the PEESE model, the variance (SE_{ij}^2) is used instead of the standard error in the RVE meta-regression model.

$$g_{ij} = \alpha_{PEESE} + SE_{ij}^2\gamma + v_j + \varepsilon_{ij} \qquad (4)$$

The traditional FAT-PET-PEESE approach has been noted to increase Type I error rates when effect sizes are dependent. To address this, we use Robust Variance Estimation (RVE), which Rodgers and Pustejovsky (2021) have recently validated as an adequate method. While we explored other publication bias correction methods, such as trim-and-fill or parametric selection models, recent research, including Kvarven et al. (2020), indicates that the PET-PEESE models generally perform better than these alternatives in most scenarios.

Figure C1: Funnel plot for type 1 studies

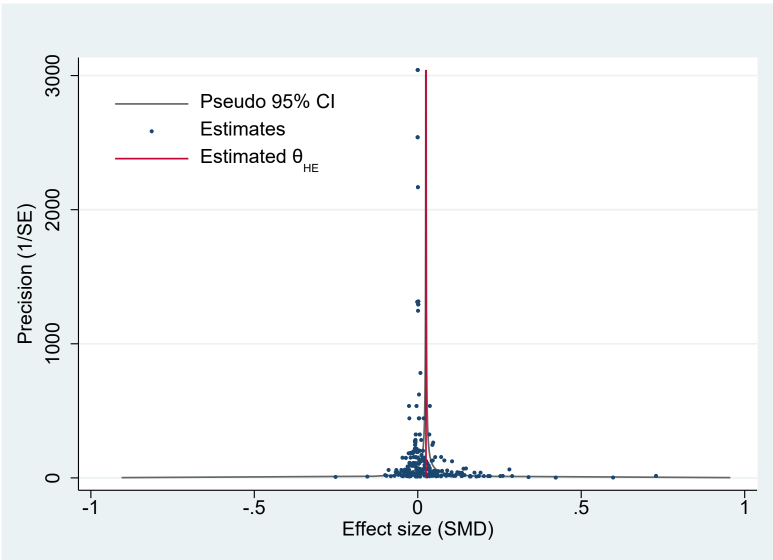


*Figure notes: This figure presents a funnel plot of effect sizes of type 1 studies plotted against their precision, defined as the inverse of their standard error. The red line indicates the unweighted mean effect size (g = 0.023***, see Table C1). The blue bell-shaped curve depicts the boundaries of the 5% significance level (i.e., 95% CI).*

Table C1: RVE univariate meta-regression for type 1 studies

	All	Employment	Earnings	Experimental & without unconfound.	With unconfound.
Average effect	0.023***	0.032***	0.014**	0.013**	0.137*
	(0.007)	(0.011)	(0.006)	(0.006)	(0.078)
	[0.009, 0.037]	[0.012, 0.053]	[0.002, 0.026]	[0.003, 0.024]	[-0.016, 0.289]
Estimates	317	98	219	270	47
Cluster	40	25	35	29	11
Studies	39	23	36	29	10

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

Figure C2: Funnel plot for type 2 studies

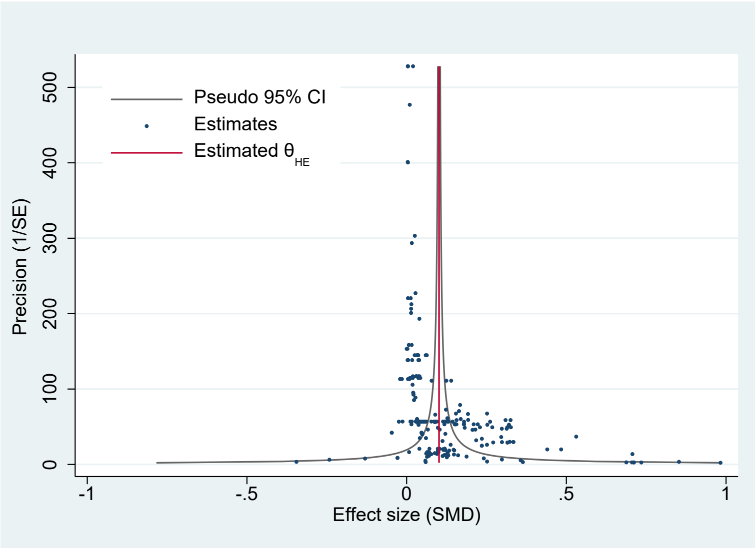


Figure notes: This figure presents a funnel plot of effect sizes of type 1 studies plotted against their precision, defined as the inverse of their standard error. The red line indicates the unweighted mean effect size (g = 0.075, see Table C2). The blue bell-shaped curve depicts the boundaries of the 5% significance level (i.e., 95% CI).

Table C2: Univariate Meta-Regression for type 2 studies

	All	Employment	Earnings	Experimental & without unconfound.
Average effect	0.075***	0.075***	0.071***	0.077***
	(0.013)	(0.017)	(0.014)	(0.014)
	[0.049, 0.100]	[0.042, 0.107]	[0.044, 0.097]	[0.049, 0.104]
Estimates	243	62	181	214
Cluster	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

D. Robustness

D.1 The impact of VET vs. general education

D.1.1 Multivariate RVE estimations for employment- and earnings-related outcomes

Table D1: Multivariate Meta-Regression with Publication Bias, employment-related outcomes

	1	2	3	4	5	6
Var (SMD)	-0.911	-0.960	-1.294	-1.783	-1.157	-2.413
	(5.205)	(5.224)	(5.710)	(5.396)	(5.677)	(5.836)
Exper. & Without unconfound.	-0.058	-0.047	-0.054	-0.049	-0.047	-0.045
Baseline: with unconfound.	(0.037)	(0.038)	(0.046)	(0.040)	(0.044)	(0.045)
Female participants		-0.039**			-0.031	-0.022
Baseline: Pooling fem. & male		(0.017)			(0.019)	(0.020)
Male participants		-0.023			-0.025	-0.017
Baseline: Pooling fem. & male		(0.020)			(0.017)	(0.018)
Average age		0.000			0.001	0.001
		(0.001)			(0.001)	(0.001)
Program exit: 6-10 years ago			-0.017		-0.020	-0.059
Baseline: up to 5 years			(0.026)		(0.025)	(0.036)
Program exit: >10 years ago			-0.020		-0.029	-0.035
Baseline: up to 5 years			(0.027)		(0.030)	(0.036)
Program durat.: 13-24 months				0.068***		0.037
Baseline: up to 12 months				(0.025)		(0.034)
Program durat.: 25-36 months				0.049***		0.038*
Baseline: up to 12 months				(0.014)		(0.023)
Program at tertiary level				0.045		0.061**
Baseline: at secondary level				(0.032)		(0.027)
Constant	0.084**	0.077	0.094*	0.014	0.066	0.031
	(0.037)	(0.053)	(0.051)	(0.040)	(0.057)	(0.065)
Obs	98	98	98	88	98	88
Cluster	25	25	25	22	25	22
Studies	23	23	23	22	23	22

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

Table D2: Multivariate Meta-Regression with Publication Bias, earnings related outcomes

	1	2	3	4	5	6
Var (SMD)	3.566	3.033*	3.510	3.251	3.007*	2.647*
	(2.278)	(1.834)	(2.245)	(2.078)	(1.813)	(1.595)
Exper. & Without unconfound.	-0.052	-0.058	-0.058	-0.072*	-0.065	-0.086
Baseline: with unconfound.	(0.042)	(0.052)	(0.044)	(0.038)	(0.052)	(0.057)
Female participants		-0.003			0.006	0.004
Baseline: Pooling fem. & male		(0.011)			(0.012)	(0.016)
Male participants		0.052			0.061	0.055
Baseline: Pooling fem. & male		(0.043)			(0.043)	(0.048)
Average age		-0.002**			-0.001	-0.001
		(0.001)			(0.001)	(0.001)
Program exit: 6-10 years ago			-0.010		0.002	-0.008
Baseline: up to 5 years			(0.012)		(0.014)	(0.013)
Program exit: >10 years ago			-0.024		-0.017	-0.005
Baseline: up to 5 years			(0.015)		(0.017)	(0.018)
Program durat.: 13-24 months				-		-0.059
				0.049**		
Baseline: up to 12 months				(0.023)		(0.037)
Program durat.: 25-36 months				-0.002		-0.001
Baseline: up to 12 months				(0.017)		(0.012)
Program at tertiary level				0.003		0.006
Baseline: at secondary level				(0.069)		(0.092)
Constant	0.060	0.116**	0.078*	0.089**	0.112**	0.133**
	(0.041)	(0.052)	(0.042)	(0.041)	(0.054)	(0.059)
Obs	219	219	219	218	219	218
Cluster	35	35	35	34	35	34
Studies	36	36	36	35	36	35

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

D.1.2 Unrestricted Weighted Least Squares

Table D3: Univariate Meta-Regression (UWLS)

	All	Employment	Earnings	Experimental & Without unconfound.	With unconfound.
Average effect	0.000***	0.002***	0.000***	0.000***	0.081***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.014)
	[0.000, 0.000]	[0.001, 0.004]	[0.000, 0.000]	[0.000, 0.000]	[0.050, 0.111]
Estimates	317	98	219	270	47
Clusters	40	25	35	29	11
Studies	39	23	36	29	10

Table notes: *p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.

Table D4: Univariate Meta-Regression with Publication Bias (PET) (UWLS)

	All	Employment	Earnings	Experimental & Without unconfound.	With unconfound.
Publ. Bias	0.966**	1.430**	0.734*	0.745*	0.604
	(0.382)	(0.556)	(0.412)	(0.415)	(0.689)
	[0.193, 1.738]	[0.283, 2.577]	[-0.103, 1.571]	[-0.106, 1.596]	[-0.932, 2.140]
Average effect	0.000*	0.000	0.000**	0.000**	0.065**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.022)
	[-0.000, 0.000]	[-0.000, 0.001]	[0.000, 0.000]	[0.000, 0.000]	[0.015, 0.114]
Estimates	317	98	219	270	47
Clusters	40	25	35	29	11
Studies	39	23	36	29	10

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias=SMD SE.

Table D5: Univariate Meta-Regression with Publication Bias (PEESE) (UWLS)

	All	Employment	Earnings	Experimental & Without unconfound.	With unconfound.
Publ. Bias	13.167**	17.612**	11.605**	16.575**	3.551
	(4.926)	(7.349)	(5.340)	(6.826)	(2.483)
	[3.202, 23.131]	[2.444, 32.779]	[0.753, 22.457]	[2.593, 30.556]	[-1.981, 9.084]
Average effect	0.000***	0.002***	0.000***	0.000***	0.077***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.014)
	[0.000, 0.000]	[0.001, 0.003]	[0.000, 0.000]	[0.000, 0.000]	[0.046, 0.108]
Estimates	317	98	219	270	47
Clusters	40	25	35	29	11
Studies	39	23	36	29	10

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias=SMD VAR.

Table D6: Univariate Meta-Regression with Publication Bias (PEESE) (UWLS)

	Employment	Earnings
Publ. Bias	12.358*	28.529
	(6.853)	(26.999)
	[-1.987, 26.702]	[-26.690, 83.749]
Average effect	0.012***	-0.000***
	(0.002)	(0.000)
	[0.008, 0.017]	[-0.000, -0.000]
Estimates	63	160
Clusters	20	30
Studies	17	29

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.
Variable to account for publication bias= SMD VAR.

Table D7: Multivariate Meta-Regression with Publication Bias, pooled outcomes (UWLS)

	1	2	3	4	5	6
Var (SMD)	1.357***	0.909	1.013*	0.870*	0.936	0.375
	(0.396)	(0.591)	(0.524)	(0.491)	(0.598)	(0.585)
Exper. & Without unconfound.	-0.121	-0.167*	-0.117	-0.161	-0.155*	-0.170**
Baseline: with unconfound.	(0.105)	(0.091)	(0.091)	(0.100)	(0.089)	(0.076)
Female participants		0.013			0.017	-0.070
Baseline: Pooling fem. & male		(0.048)			(0.043)	(0.057)
Male participants		0.117**			0.114**	0.021
Baseline: Pooling fem. & male		(0.046)			(0.046)	(0.060)
Average age		-0.008*			-0.005	-0.005
		(0.004)			(0.006)	(0.005)
Program exit: 6-10 years ago			-0.109*		-0.073	-0.141***
Baseline: up to 5 years			(0.055)		(0.062)	(0.042)
Program exit: >10 years ago			-0.137*		-0.062	-0.058
Baseline: up to 5 years			(0.072)		(0.085)	(0.061)
Program durat.: 13-24 months				-0.115**		-0.094
Baseline: up to 12 months				(0.047)		(0.065)
Program durat.: 25-36 months				0.100**		0.125*
Baseline: up to 12 months				(0.044)		(0.065)
Program at tertiary level				-0.133		-0.112
Baseline: at secondary level				(0.097)		(0.100)
Constant	0.178*	0.433*	0.248**	0.178*	0.370	0.352*
	(0.090)	(0.230)	(0.121)	(0.103)	(0.234)	(0.194)
Obs	317	317	317	306	317	306
Cluster	40	40	40	37	40	37
Studies	39	39	39	38	39	38

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR.

D.1.3 Ordinary Least Squares

Table D8: Univariate Meta-Regression (OLS)

	All	Employment	Earnings	Experimental & Without unconfound.	With unconfound.
Average effect	0.035**	0.035**	0.035**	0.021	0.114***
	(0.013)	(0.013)	(0.016)	(0.013)	(0.028)
	[0.008, 0.062]	[0.008, 0.063]	[0.003, 0.067]	[-0.006, 0.048]	[0.053, 0.176]
Estimates	317	98	219	270	47
Clusters	40	25	35	29	11
Studies	39	23	36	29	10

*Table notes: *p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.*

Table D9: Univariate Meta-Regression with Publication Bias (PET) (OLS)

	All	Employment	Earnings	Experimental & Without unconfound.	With unconfound.
Publ. Bias	1.114***	0.405	1.234***	0.886	0.982***
	(0.267)	(0.460)	(0.293)	(0.687)	(0.285)
	[0.575, 1.653]	[-0.545, 1.355]	[0.638, 1.830]	[-0.522, 2.294]	[0.346, 1.618]
Average effect	0.000	0.022	-0.002	-0.000	0.047
	(0.008)	(0.014)	(0.008)	(0.010)	(0.028)
	[-0.015, 0.016]	[-0.007, 0.050]	[-0.019, 0.015]	[-0.022, 0.021]	[-0.015, 0.109]
Estimates	317	98	219	270	47
Clusters	40	25	35	29	11
Studies	39	23	36	29	10

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias=SMD SE.*

Table D10: Univariate Meta-Regression with Publication Bias (PEESE) (OLS)

	All	Employment	Earnings	Experimental & Without unconfound.	With unconfound.
Publ. Bias	2.643***	3.453	2.638***	7.856	2.023**
	(0.889)	(4.774)	(0.887)	(6.534)	(0.722)
	[0.844, 4.441]	[-6.401, 13.307]	[0.835, 4.441]	[-5.528, 21.240]	[0.414, 3.632]
Average effect	0.028**	0.029**	0.027*	0.011	0.093***
	(0.012)	(0.012)	(0.014)	(0.007)	(0.027)
	[0.003, 0.052]	[0.003, 0.054]	[-0.002, 0.055]	[-0.004, 0.026]	[0.033, 0.153]
Estimates	317	98	219	270	47
Clusters	40	25	35	29	11
Studies	39	23	36	29	10

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.
Variable to account for publication bias= SMD VAR.

Table D11: Univariate Meta-Regression with Publication Bias (PEESE) (OLS)

	Employment	Earnings
Publ. Bias	-0.555	-0.005
	(3.726)	(1.135)
	[-8.354, 7.243]	[-2.327, 2.317]
Average effect	0.040**	0.050
	(0.017)	(0.045)
	[0.005, 0.075]	[-0.041, 0.141]
Estimates	63	160
Clusters	20	30
Studies	17	29

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets = Confidence Interval. Variable to account for publication bias= SMD VAR.

Table D12: Multivariate Meta-Regression with Publication Bias, pooled outcomes (OLS)

	1	2	3	4	5	6
Var (SMD)	2.217***	2.103***	2.055***	2.128**	2.040***	1.942**
	(0.819)	(0.776)	(0.756)	(0.796)	(0.745)	(0.718)
Exper. & Without unconfound.	-0.073**	-0.078**	-0.074**	-0.085***	-0.077**	-0.085***
Baseline: with unconfound.	(0.030)	(0.032)	(0.034)	(0.024)	(0.032)	(0.026)
Female participants		0.002			0.008	0.008
Baseline: Pooling fem. & male		(0.015)			(0.014)	(0.013)
Male participants		0.029			0.033*	0.028
Baseline: Pooling fem. & male		(0.020)			(0.019)	(0.019)
Average age		-0.002			-0.000	-0.000
		(0.001)			(0.001)	(0.001)
Program exit: 6-10 years ago			-0.031		-0.037	-0.057*
Baseline: up to 5 years			(0.025)		(0.028)	(0.028)
Program exit: >10 years ago			-0.038		-0.040	-0.040
Baseline: up to 5 years			(0.025)		(0.029)	(0.030)
Program durat.: 13-24 months				-0.041*		-0.026
Baseline: up to 12 months				(0.021)		(0.021)
Program durat.: 25-36 months				0.012		0.027
Baseline: up to 12 months				(0.023)		(0.020)
Program at tertiary level				0.028		0.043
Baseline: at secondary level				(0.032)		(0.029)
Constant	0.091***	0.142***	0.116***	0.099***	0.121***	0.110***
	(0.026)	(0.047)	(0.036)	(0.029)	(0.043)	(0.036)
Obs	317	317	317	306	317	306
Cluster	40	40	40	37	40	37
Studies	39	39	39	38	39	38

Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR.

D.2 The impact of VET vs. an unrestricted control group

D.2.1 Multivariate RVE estimations for employment- and earnings-related outcomes

Table D13: Multivariate Meta-Regression with Publication Bias, employment-related outcomes

	1	2	3	4	5	6
Var (SMD)	4.036***	3.874***	3.935***	3.734***	3.895***	3.109
	(0.702)	(0.935)	(0.658)	(0.704)	(0.888)	(2.078)
Exper. & Without unconfound.	0.028	0.013	0.032	0.029	-0.007	-0.048
Baseline: with unconfound.	(0.019)	(0.015)	(0.025)	(0.043)	(0.042)	(0.070)
Female participants		-0.039			-0.046	-0.088
Baseline: Pooling fem. & male		(0.059)			(0.060)	(0.089)
Male participants		-0.057			-0.042	-0.067
Baseline: Pooling fem. & male		(0.048)			(0.054)	(0.105)
Average age		-0.001			-0.001	-0.003
		(0.003)			(0.004)	(0.006)
Program exit: 6-10 years ago			-0.001		0.034	0.061
Baseline: up to 5 years			(0.031)		(0.047)	(0.069)
Program exit: >10 years ago			-0.065***		-0.030	0.026
Baseline: up to 5 years			(0.022)		(0.053)	(0.115)
Program durat.: *0-12 months*				-0.072		-0.087
Baseline: 25-36 months				(0.062)		(0.225)
Program durat.: 13-24 months				-0.046		-0.109
Baseline: 25-36 months				(0.057)		(0.234)
Program at tertiary level				0.058		0.124
				(0.043)		(0.135)
Constant	0.043***	0.109	0.055**	0.044***	0.123	0.232
	(0.006)	(0.163)	(0.022)	(0.006)	(0.154)	(0.251)
Obs	62	56	62	62	56	56
Cluster	19	17	19	19	17	17
Studies	13	12	13	13	12	12

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

Table D14: Multivariate Meta-Regression with Publication Bias, earnings related outcomes

	1	2	3	4	5
Var (SMD)	4.130	3.858	4.069	3.629	3.826
	(2.740)	(2.830)	(2.618)	(2.830)	(2.723)
Exper. & Without unconfound.	0.032*	0.036	0.038*	-0.030	0.039*
Baseline: with unconfound.	(0.018)	(0.023)	(0.020)	(0.039)	(0.022)
Female participants		0.011			0.010
Baseline: Pooling fem. & male		(0.029)			(0.028)
Male participants		0.009			0.018
Baseline: Pooling fem. & male		(0.034)			(0.037)
Average age		0.000			0.001
		(0.002)			(0.002)
Program exit: 6-10 years ago			0.020		0.018
Baseline: up to 5 years			(0.030)		(0.032)
Program exit: >10 years ago			-0.019		-0.029
Baseline: up to 5 years			(0.038)		(0.045)
Program durat.: *0-12 months*				0.023	
Baseline: 25-36 months				(0.044)	
Program durat.: 13-24 months				0.065*	
Baseline: 25-36 months				(0.039)	
Program at tertiary level				0.015	
				(0.042)	
Constant	0.038***	0.029	0.028	0.036***	0.004
	(0.010)	(0.051)	(0.021)	(0.013)	(0.048)
Obs	181	179	181	175	179
Cluster	31	30	31	30	30
Studies	26	25	26	25	25

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Setting Rho at 0.8. Variable to account for publication bias=SMD SE. Model estimated by RVE, setting Rho at 0.8. Dependent variable: SMD.*

D.2.2 Unrestricted Weighted Least Squares

Table D15: Univariate Meta-Regression (UWLS)

	All	Employment	Earnings	Experimental & Without unconfound.
Average effect	0.022**	0.039***	0.020**	0.022**
	(0.008)	(0.013)	(0.008)	(0.008)
	[0.005, 0.038]	[0.012, 0.067]	[0.004, 0.036]	[0.005, 0.038]
Estimates	243	62	181	214
Clusters	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.*

Table D16: Univariate Meta-Regression with Publication Bias (PET) (UWLS)

	All	Employment	Earnings	Experimental & Without unconfound.
Publ. Bias	4.511***	2.612***	5.044***	5.204***
	(0.751)	(0.831)	(0.793)	(0.683)
	[2.981, 6.040]	[0.866, 4.359]	[3.426, 6.663]	[3.808, 6.601]
Average effect	-0.001	0.014	-0.003	-0.004*
	(0.003)	(0.013)	(0.002)	(0.002)
	[-0.007, 0.005]	[-0.014, 0.041]	[-0.008, 0.002]	[-0.008, 0.001]
Estimates	243	62	181	214
Clusters	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD SE.*

Table D17: Univariate Meta-Regression with Publication Bias (PEESE) (UWLS)

	All	Employment	Earnings	Experimental & Without unconfound.
Publ. Bias	13.973***	6.716***	18.810*	16.396**
	(4.804)	(2.076)	(9.312)	(6.232)
	[4.188, 23.757]	[2.354, 11.077]	[-0.208, 37.829]	[3.649, 29.143]
Average effect	0.021**	0.038***	0.019**	0.021**
	(0.008)	(0.013)	(0.007)	(0.008)
	[0.005, 0.037]	[0.011, 0.065]	[0.004, 0.034]	[0.005, 0.037]
Estimates	243	62	181	214
Clusters	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR.*

Table D18: Univariate Meta-Regression with Publication Bias (PEESE) (UWLS)

	Employment	Earnings
Publ. Bias	-1.674	-11.488
	(2.529)	(33.194)
	[-6.988, 3.640]	[-80.154, 57.178]
Average effect	0.136***	0.270***
	(0.014)	(0.000)
	[0.106, 0.167]	[0.270, 0.270]
Estimates	66	122
Clusters	19	24
Studies	13	20

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets = Confidence Interval. Variable to account for publication bias= SMD VAR.*

Table D19: Multivariate Meta-Regression with Publication Bias, pooled outcomes (UWLS)

	1	2	3	4	5	6
Var (SMD)	3.625***	2.669***	3.309***	3.756***	2.499***	2.475***
	(0.536)	(0.407)	(1.010)	(0.729)	(0.425)	(0.403)
Exper. & Without unconfound.	0.425***	0.445***	0.197**	0.286***	0.279***	0.301**
Baseline: with unconfound.	(0.060)	(0.079)	(0.080)	(0.026)	(0.049)	(0.126)
Female participants		0.100			0.005	0.006
Baseline: Pooling fem. & male		(0.075)			(0.063)	(0.065)
Male participants		-0.173			-0.238	-0.241
Baseline: Pooling fem. & male		(0.279)			(0.217)	(0.227)
Average age		-0.005			0.005	0.007
		(0.005)			(0.007)	(0.005)
Program exit: 6-10 years ago			-0.079		-0.216**	-0.221***
Baseline: up to 5 years			(0.126)		(0.087)	(0.072)
Program exit: >10 years ago			-0.272***		-0.304**	-0.314***
Baseline: up to 5 years			(0.053)		(0.136)	(0.108)
Program durat.: *0-12 months*				0.017		-0.004
Baseline: 25-36 months				(0.035)		(0.098)
Program durat.: 13-24 months				0.086**		-0.017
Baseline: 25-36 months				(0.034)		(0.040)
Program at tertiary level				0.076		-0.006
				(0.080)		(0.085)
Constant	-0.246***	-0.079	0.030	-0.251***	-0.068	-0.104
Obs	243	235	243	237	235	229
Cluster	33	31	33	32	31	30
Studies	27	26	27	26	26	25

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR.*

D.2.3 Ordinary Least Squares

Table D20: Univariate Meta-Regression (OLS)

	All	Employment	Earnings	Experimental & Without unconfound.
Average effect	0.119***	0.118***	0.119***	0.131***
	(0.024)	(0.036)	(0.029)	(0.026)
	[0.070, 0.169]	[0.042, 0.195]	[0.059, 0.180]	[0.078, 0.183]
Estimates	243	62	181	214
Clusters	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.*

Table D21: Univariate Meta-Regression with Publication Bias (PET) (OLS)

	All	Employment	Earnings	Experimental & Without unconfound.
Publ. Bias	1.302***	1.381***	1.297**	1.655***
	(0.391)	(0.215)	(0.611)	(0.095)
	[0.505, 2.099]	[0.929, 1.833]	[0.050, 2.544]	[1.462, 1.849]
Average effect	0.063***	0.034*	0.072***	0.067***
	(0.013)	(0.017)	(0.015)	(0.014)
	[0.036, 0.091]	[-0.002, 0.071]	[0.042, 0.102]	[0.038, 0.096]
Estimates	243	62	181	214
Clusters	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD SE.*

Table D22: Univariate Meta-Regression with Publication Bias (PEESE) (OLS)

	All	Employment	Earnings	Experimental & Without unconfound.
Publ. Bias	3.816***	3.967***	3.791***	4.303***
	(0.641)	(0.160)	(1.108)	(0.259)
	[2.511, 5.122]	[3.630, 4.304]	[1.529, 6.053]	[3.773, 4.833]
Average effect	0.092***	0.071***	0.099***	0.102***
	(0.016)	(0.018)	(0.018)	(0.016)
	[0.060, 0.125]	[0.034, 0.108]	[0.063, 0.135]	[0.069, 0.135]
Estimates	243	62	181	214
Clusters	33	19	31	30
Studies	27	13	26	24

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR.*

Table D23: Univariate Meta-Regression with Publication Bias (PEESE) (OLS)

	Employment	Earnings
Publ. Bias	1.322***	-6.289***
	(0.259)	(1.706)
	[0.778, 1.866]	[-9.818, -2.761]
Average effect	0.103***	0.259***
	(0.018)	(0.069)
	[0.064, 0.142]	[0.117, 0.402]
Estimates	66	122
Clusters	19	24
Studies	13	20

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval.*
Variable to account for publication bias= SMD VAR.

Table D24: Multivariate Meta-Regression with Publication Bias, pooled outcomes (OLS)

	1	2	3	4	5	6
Var (SMD)	3.860***	3.836***	3.810***	3.880***	3.854***	3.962***
	(0.527)	(0.526)	(0.476)	(0.544)	(0.453)	(0.470)
Exper. & Without unconfound.	0.109***	0.101***	0.073***	0.039	0.072***	0.016
Baseline: with unconfound.	(0.034)	(0.034)	(0.011)	(0.037)	(0.013)	(0.033)
Female participants		-0.006			-0.004	-0.018
Baseline: Pooling fem. & male		(0.032)			(0.031)	(0.031)
Male participants		0.007			0.017	0.006
Baseline: Pooling fem. & male		(0.044)			(0.044)	(0.043)
Average age		-0.002			0.001	0.002
		(0.002)			(0.002)	(0.002)
Program exit: 6-10 years ago			0.001		-0.002	-0.020
Baseline: up to 5 years			(0.030)		(0.033)	(0.033)
Program exit: >10 years ago			-0.089***		-0.100***	-0.105***
Baseline: up to 5 years			(0.022)		(0.033)	(0.033)
Program durat.: *0-12 months*				-0.017		0.034
Baseline: 25-36 months				(0.032)		(0.038)
Program durat.: 13-24 months				0.053		0.062*
Baseline: 25-36 months				(0.035)		(0.035)
Program at tertiary level				0.035		0.013
				(0.032)		(0.025)
Constant	-0.004	0.059	0.045***	-0.004	0.026	-0.022
	(0.028)	(0.080)	(0.016)	(0.028)	(0.064)	(0.060)
Obs	243	235	243	237	235	229
Cluster	33	31	33	32	31	30
Studies	27	26	27	26	26	25

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR.*

D.2.4 The impact of vocational education at Community Colleges (CCs)

Table D25: RVE PEESE meta-regression: impact of vocational education at CCs

	All	Employment	Earnings	Experimental & without unconfound.
Publ. Bias	4.208***	4.117***	3.890	5.078***
	(0.898)	(0.751)	(2.657)	(0.594)
	[2.447, 5.968]	[2.645, 5.589]	[-1.317, 9.097]	[3.914, 6.242]
CC dummy	0.054***	0.055*	0.064***	0.054**
	(0.021)	(0.029)	(0.020)	(0.021)
	[0.013, 0.094]	[-0.001, 0.111]	[0.025, 0.102]	[0.012, 0.096]
Average effect remaining studies	0.039***	0.038***	0.028***	0.039***
	(0.009)	(0.012)	(0.007)	(0.011)
	[0.022, 0.057]	[0.014, 0.063]	[0.014, 0.041]	[0.017, 0.060]
Estimates (CCs)	243 (153)	62 (24)	181 (129)	214 (153)
Cluster (CCs)	33 (21)	19 (11)	31 (21)	30 (21)
Studies (CCs)	27 (17)	13 (7)	26 (17)	24 (24)

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets =Confidence Interval. Variable to account for publication bias= SMD VAR. Setting Rho at 0.8.*

Table D26: RVE PEESE regression by occupational field, using SMD as outcome variable

	All outcomes
Publ. Bias	4.217***
	(0.596)
Baseline: Human Services & Education	0.039
	(0.032)
	[-0.025, 0.102]
Communication & Information Systems	0.003
	(0.024)
	[-0.044, 0.049]
Health Sciences	0.139***
	(0.053)
	[0.035, 0.243]
Skilled Technical Sciences	0.053
	(0.047)
	[-0.038, 0.144]
Agriculture	0.147***
	(0.054)
	[0.041, 0.252]
Business, Marketing, & Management	0.016
	(0.040)
	[-0.062, 0.095]
Estimates	157
Cluster	11
Studies	11

*Table notes: * p<0.10, ** p<0.05, *** p<0.01. Rectangular brackets = Confidence Interval. Variable to account for publication bias= SMD VAR. Setting Rho at 0.8.*

ETH Zürich
Chair of Education Systems
Stampfenbachstrasse 69
8092 Zürich

www.ces.ethz.ch

Publisher: Chair of Education Systems CES
Author: ETH Zurich
Design: ETH Zurich