

Wage effects of on-the-job training: A meta-analysis

Citation for published version (APA):

Haelermans, C., & Borghans, L. (2012). Wage effects of on-the-job training: A meta-analysis. *British Journal of Industrial Relations*, 50(3), 502-528. <https://doi.org/10.1111/j.1467-8543.2012.00890.x>

Document status and date:

Published: 01/09/2012

DOI:

[10.1111/j.1467-8543.2012.00890.x](https://doi.org/10.1111/j.1467-8543.2012.00890.x)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Wage Effects of On-the-Job Training: A Meta-Analysis

Carla Haelermans and Lex Borghans

Abstract

A meta-analysis is used to study the average wage effects of on-the-job training. This study shows that the average reported wage effect of on-the-job training, corrected for publication bias, is 2.6 per cent per course. The analyses reveal a substantial heterogeneity between training courses, while wage effects reported in studies based on instrumental variables and panel estimators are substantially lower than estimates based on techniques that do not correct for selectivity issues. Appropriate methodology and the quality of the data turn out to be crucial to determine the wage returns.

1. Introduction

It has been argued frequently that on-the-job training is an important investment because of the increased importance of knowledge and skills for the labour market (OECD 2003). However, participating in training courses takes time, so the question arises what the (wage) returns to these investments are. While the return to (initial) education has been investigated extensively, much less empirical evidence is available about the wage effects of on-the-job training, mainly for statistical reasons.

The aim of this study is to investigate the wage returns to on-the-job training, taking into account the statistical complications present in existing studies. We use a meta-analysis based on 71 estimates of returns to on-the-job training from studies written and published between 1981 and 2010. The main finding of our study is that the average wage effect of on-the-job training is 2.6 per cent. Compared with the average return to education, as reported by Ashenfelter *et al.* (1999), this implies that, on average, the wage return to on-the-job training is larger than the return to education, and — assuming that the workers also bear at least the indirect costs of training — may be profitable for an individual up until the age of 61. However, this

Carla Haelermans is at Top Institute of Evidence Based Education Research, Maastricht University and Delft University of Technology. Lex Borghans is at Maastricht University.

reported return turns out to be lower in studies that use estimation techniques correcting for selectivity bias. In that case, the age until which an average training course is profitable reduces to 55. We also find a substantial heterogeneity in the wage effects of training courses. Therefore, some courses are profitable up to a higher age, whereas other courses seem not to pay even at very young ages. The implication of this heterogeneity is that knowledge about the expected wage returns of a specific training course is important for a firm when deciding whether to invest in on-the-job training. Our findings show that an appropriate statistical methodology and high-quality data are needed to estimate these wage returns.

We identify four major problems with respect to investigating the wage returns to on-the-job training. First, most studies that research the wage effects of training courses have a low power, which makes it statistically hard to obtain a reliable estimate. Because training participation is much lower than participation in education, and the average course is rather short, estimates of the wage effects based on single surveys are imprecise. Second, because of this low power, studies on the wage effects of on-the-job training are disproportionately subject to publication bias. The chances are high that analyses that find high, and therefore significant, returns are published, while articles with small estimates, and therefore insignificant coefficients, may not be accepted by journals. Third, not all studies take into account the selectivity issues with respect to who gets to participate in a training course, and who does not. Fourth, studies on the wage effects of on-the-job training are very diverse and have different characteristics, which at first sight make them hard to compare.

The meta-analysis applied in this article overcomes all the above-mentioned problems. One of the major advantages of the meta-analysis is that it helps to find an overall structure among the articles that individually might result from imprecise measures because of small samples. The danger of imprecise measures is that — due to the process of editing and refereeing — publication bias occurs. The meta-analysis can deal with this as well, by explicitly taking into account the publication bias. There are empirical approaches in the literature to tackle problems of endogeneity related to selectivity in participation, such as panel data and instrumental variable (IV) methods. A meta-analysis makes it possible to compare these different approaches with one another, and define how the choice of a specific approach affects the estimate. Furthermore, in the meta-analysis, we can also distinguish between, for example, different types of data or different circumstances that are used by individual studies. In sum, the meta-analysis increases the sample size, takes into account publication bias, and makes it possible to compare the seemingly incomparable studies.

The remainder of the article is as follows: first, in Section 2, we discuss some of the previous research and reviews related to meta-analysis, human capital theory, and on-the-job training. We then outline the framework of this study. In Section 3, we describe the database, and explain the meta-analysis and the meta-regression analysis methodology. Because these

methods are less common in the field of industrial relations, we provide the basic concepts of these research methods. In Section 4, we present the results of the meta-analysis, followed by the meta-regression analysis in Section 5. Finally, in Section 6, we end with some concluding remarks.

2. Previous research and reviews

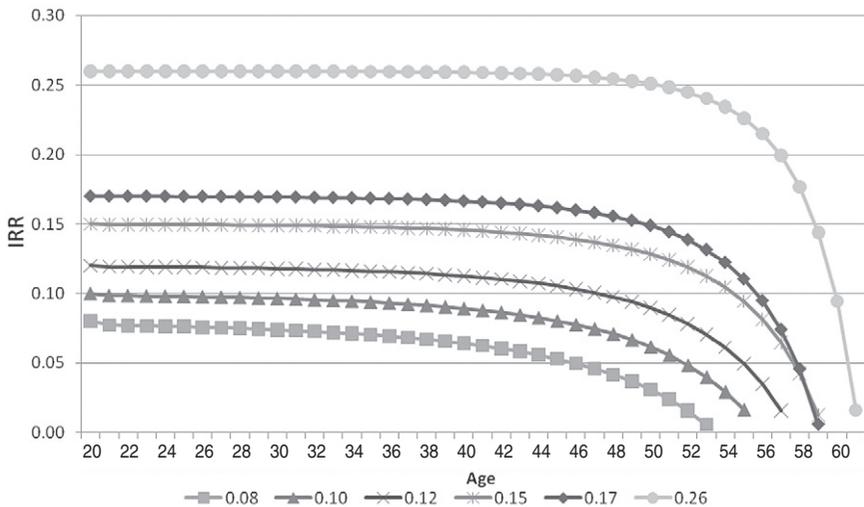
Two studies have investigated the effectiveness of training, using a meta-analysis (Greenberg *et al.* 2003, 2006). Greenberg *et al.* (2003) use a mixed-effects meta-regression model to estimate the overall effect of government-funded training programmes, and find that the earnings effect is largest for women and that the effect persists some years after the training is completed. Greenberg *et al.* (2006), applying the same dataset, compare studies that have used observational methods with those using experimental methods. The weighted regression is estimated with the mixed-effects model. They conclude that the results of the two approaches do not differ much.

In this article, the wage effects of on-the-job training are also investigated using meta-analysis. Becker (1964) developed the idea that education and training can be regarded as investments. Like investments in physical capital, investments can also be made in people by increasing their productivity through learning. Because learning takes time, the main costs of these investments are the foregone earnings. The relative increase in wages related to one year of education is therefore referred to as the 'return to education' (Mincer 1958). Ben-Porath (1967) investigated the optimal investment pattern of human capital over the life cycle. In an optimal setting, people invest in human capital full-time early on in life. Later, they reduce the time spent on education to reap the benefits in the labour market. So while education will be beneficial to all people early in life, the efficiency of training adults will depend on specific circumstances and the content of the course. For the person involved, education and training already obtained will be relevant for the value of additional training. Especially for people who, for whatever reason, have missed crucial earlier investments, training might be beneficial. Circumstances that can influence the value of a training course could be technological developments or shifts in the demand for labour across industries.

The literature distinguishes between training for people who are employed and training to bring people back to work. Most of the literature focuses on employment programmes (Heckman *et al.* 1999). A large share of the literature about on-the-job training focuses on the question *why* firms are willing to invest in general training. Estimates of the wage effect of on-the-job training are relatively scarce. The contribution of this article is that it brings together this evidence. By comparing different studies, it is possible to investigate both the average wage effect of training and the variation in the wage effect. Furthermore, it is possible to investigate whether estimates are also related to the methodology used when looking at the wage effect of on-the-job training. The articles that are analysed in this meta-analysis typically

measure the relative increase in wages as a result of training. Despite the similarity with the return to education literature, it is important to note that this measure is not a return to training. First, not all the returns will be reflected in the wage. Especially in case of firm-specific human capital, higher productivity does not necessarily lead to higher wages. But also the reverse might be true. Employers could pay for a training course that increases the wage of a worker. In that case, from the individual point of view, the investment could be beneficial even though the costs exceed the returns. Second, the average training course takes much less than one year. Opportunity costs are therefore much lower. Third, workers who take part in on-the-job training are older than students in education because they are already in the workforce. The typical measure for the rate of return to education assumes that forgone earnings are the only costs of education, and assumes an increase in earnings due to this year of education for an infinite time horizon. In that case, the internal rate of return (IRR) of this investment of one year of education equals the relative increase in wages. The investment in education gives the same annual return as this worker would have obtained if he had put one year of income in a bank account with a discount rate equal to this wage increase. For young workers, the assumption of infinite horizons might be reasonable, but for older workers, this will overestimate their returns because these workers have less time left in their working life to earn back the investment. This is illustrated in Figure 1. The figure provides simulated values of the rate of return (IRR) for investments in education or training, varying the age for investments with a wage effect of 8, 10, 12, 15, 17 and 26 per cent. From the age of 53 to 62 (depending on the IRR), this rate

FIGURE 1
Relationship between the Internal Rate of Return and the Wage Effect of Training,
Depending on the Number of Years before Retirement.



of return is negative, indicating that, even with a 0 per cent discount rate, these investments do not pay off. Between the ages of 30 and 40 (and younger), the rate of return is close to the relative wage effect, but, for people who are in the middle of their career, the rate of return of an equal-sized wage effect is substantially lower. A relative wage increase per year of training that is equal to the relative wage increase of a year of schooling implies that this investment in training is less beneficial than an investment in education. For individual workers, it can still be worthwhile making the investment, as it is too late for them to invest earlier in life, but from a societal point of view, these investments might be inefficient.

Theoretical work about on-the-job training has focused on the division of the cost and return to training (for an overview, see Frazis and Loewenstein 2006). The main issue is that on-the-job training can be firm specific. That means that the skills obtained only have added value for the worker in the firm concerned. Therefore, an increase in the workers' productivity does not necessarily lead to an increase in wages. Loewenstein and Spletzer (1999) find empirical evidence for this theoretical distinction between general and firm-specific training. Unfortunately, most studies on returns to training do not report whether the training was general or firm specific.

As a consequence, in empirical work about the effects of on-the-job training, the distinction between productivity and wage effects is relevant. Empirically, this distinction is hard to make because productivity is difficult to measure. Recently, some studies have used productivity measures such as value added (Almeida and Carneiro 2006; Conti 2005; Dearden *et al.* 2006; Schonewille 2001), gross revenue per worker, sales (Barret and O'Connell 2001), and employer scales of productivity of workers (Barron *et al.* 1998; Groot 1999). The number of studies and the diversity in productivity measures do not allow a meta-analysis on the productivity effects of training. In this study, we therefore restrict the analysis to the wages effect of on-the-job training.

One of the main concerns in the literature on the return to education is the potential bias that could arise from selectivity in educational decisions (Griliches 1977). The choice for IV or natural experiments over ordinary least squares (OLS) is therefore a major issue in this literature. Similar approaches to estimate the wage effects of on-the-job training are rare, and the instruments used are often problematic. Kuckulenz and Zwick (2003), for example, use organization size; Leuven and Oosterbeek (2008) base their estimates on workers who coincidentally did not attend a planned course; and Leuven and Oosterbeek (2004) use tax discontinuities.¹ Loewenstein and Spletzer (1997) take expected job change as an instrument. In general, also fixed-effect panel data approaches can be used to establish a causal relationship. However, in the case of the return to education, this is not feasible because we need wages before and after the education. For the effects of on-the-job training, fixed-effects approaches are an option. Three studies in our database have used fixed effects (FE): namely Krueger and Rouse (1998), Melero (2004), and Xiao (2002).

The research framework of this study compares on-the-job training with education. Both education and on-the-job training are time-consuming and costly. The question is whether it is worth investing in these types of education. In this study, we compare the wage effects of education with the wage effects of on-the-job training. The research on daily time spending from Statistics Netherlands (2000) shows that 16-year-old pupils spent, on average, 5,657 hours per day on schooling, seven days a week, holidays not included. Based on 40 weeks of school per year, this puts a year of education equal to about 1600 hours of schooling. The data on duration of training in the International Adult Literacy Study (OECD 2000) show that on-the-job training programmes take, on average, 137 hours per year. A meta-study by Ashenfelter *et al.* (1999) reports that the average returns to education are 8.1 per cent. The standard deviation of the returns between studies is three percentage points. Therefore, if returns to training are about equal to returns to education, one could expect the wage effect of a training course to equal approximately 0.7 per cent. The number of 0.7 seems to be a reasonable reference point for the wage effects of on-the-job training. We place the wage effect of on-the-job training in the perspective of the figure of the IRR (Figure 1), which leads to the maximum age that justifies investment in on-the-job training.

3. Meta-analysis and meta-regression

We use two of the main meta-approaches: meta-analysis and meta-regression analysis. We use meta-analysis to estimate the average effect size, whereas meta-regression analysis is used to show how differences between the studies can be explained. In this section, we first discuss how we constructed the database and then describe the methodology of meta-analysis. Next, we explain the problem of publication bias and present the meta-regression analysis.

The Database

We collected all published articles from the *Economic Literature Index*, *JSTOR*, *Web of Science*, *Interscience* and *Science Direct*. In addition, we collected unpublished articles from the *Institute for the Study of Labour* and the *National Bureau of Economic Research*. The keywords used during this search process are the following: training, effects of training, training and wages, and on-the-job-training. To select studies, the following criteria were applied:

1. The study includes an empirical analysis of on-the-job training.
2. The study computes the effect of training on wages.
3. The study uses training as a dummy variable, thereby ensuring the existence of both a treatment and a control group.²

4. The number of observations should be given, or it should be possible to calculate them.
5. The study reports a difference effect between the treatment and the control group, or effects for both groups, and the subsequent standard deviation, standard error, *t*-statistic, *p*-value or *F*-statistic.
6. The study is written in English.

The database was coded by both authors. As the coding scheme was quite straightforward, there were hardly any conflicts on the coding decisions. In cases where conflicts arose, a third party was consulted. The effect size and its standard error are the main variables needed for a meta-analysis. The effect size is the regression estimate of the treatment compared with the reference group as reported in the studies. Because all studies use the logarithm of the wage, our effect size represents the percentage difference in the effect of training between the treatment and the control group. Note that studies on training for the unemployed are not part of our analysis. Some studies report more than one effect size: for example, for male and female separately, or for different control groups. We only include more than one effect size per study when there is no overlap between the observations used for the effect sizes, meaning that an observation can only be used for one of the effect sizes. If more than one effect size is estimated in one article and there appears to be an overlap, the estimate that is marked by the author of that particular study as the most important one is used for our meta-analysis, as we assume that authors who present only one effect size have made their decision based on the same argument. We furthermore exclude studies in which authors use a dataset they have used before in the same context, to make sure all studies are statistically independent. The deadline for entries in this study was January 2011. Because of the above-mentioned selection criteria, many studies that were found using the above-mentioned keywords had to be excluded. Most studies that were excluded did not include training as a dummy variable in their estimation but used, for example, hours of training or number of training courses per year. Because the analysis of the effect of, for example, an extra hour of training is hard to compare with having had some training, we decided to exclude the studies that do not use a dummy variable. Last, many studies were excluded because they did not distinguish between a treatment group and a control group. The process of excluding studies was an ongoing process from the start of the database search. Many studies were already selected based on their abstracts and keywords, whereas other studies were excluded based on the above-mentioned criteria. Therefore, it is impossible to say exactly how many studies were excluded in the process.

In total, 71 estimates from 38 studies were selected for this meta-analysis. The complete database and the corresponding coding scheme are available upon request from the corresponding author. In the database, we separate estimates based on the author of the study, the country to which the dataset relates, the gender of the individuals, the type of wage measurement to which the effects of training are related (yearly wage, monthly wage, hourly wage or

other), the estimation method the study has used (IV, OLS or FE), the way of precision,³ reported in the publication, with which the estimate is calculated (standard errors or *t*-statistics), article status (published or not), the first year of data used, and the total number of observations. Unfortunately, we cannot make a distinction between general and specific training, or include information on the duration of the training, as most studies do not report this information. This information will be hidden in the study heterogeneity. Furthermore, the database includes the effect size of on-the-job training on wages and the corresponding standard error of that effect size. Studies report either standard errors or *t*-values. We transformed these *t*-values into standard errors of the parameters. The studies used for the analysis were published between 1981 and 2010, the first year of the data used ranging from 1952 to 2005. The total number of observations varies from 168 to 34,900. Most of the studies use individuals of working age for their estimation, using OLS. Furthermore, six studies estimated a negative effect size.

Meta-analysis

In the basic model of meta-analysis, the assumption is that the true effect size is the same for every study. Differences between estimated effect sizes only occur because of sampling error. In this case, a fixed-effects model can be used to estimate the weighted mean effect size and its standard error. Therefore, we use the following:

$$\overline{ES} = \frac{\sum(w_i ES_i)}{\sum w_i} \tag{1}$$

$$SE_{\overline{ES}} = \sqrt{\frac{1}{\sum w_i}} \tag{2}$$

where \overline{ES} is the weighted mean effect size, and $SE_{\overline{ES}}$ is the standard error of the mean; ES_i is the individual effect size for $i = 1$ to k (where k is the number of effect sizes); and w_i is the individual weight for ES_i . The inverse variance (w_i) is used to determine the weight of a study in the meta-analysis and is given by the following:

$$v_i = SE_i^2 \tag{3}$$

$$w_i = \frac{1}{SE_i^2} \tag{4}$$

where v_i is the subject level sampling error of each study. The assumption that the effect size distribution is random requires homogeneity. To test for homogeneity, we calculate the *Q*-statistic, according to the following:

$$Q = \sum w_i (ES_i - \overline{ES})^2 \quad (5)$$

The Q -statistic is distributed as a chi-square, with $k - 1$ degrees of freedom (Hedges and Olskin 1985), where k is the number of studies included. When Q is found to be significant, the null hypothesis of homogeneity is rejected. If this is the case, the FE model is not the best estimation method, as it ignores the extra variability due to differences between studies in addition to the sampling error.

One way of solving this is using the random-effects model for the estimation of the weighted mean effect size. The random-effects method assumes that the variation not only exists at the subject level but also exists at the study level. Incorporating both sampling error levels gives us a different inverse variance weight and therefore a different effect size. For the random-effects method, a new value of the variance has to be calculated, which is the sum of the subject-level sampling error (v_i), defined in equation (3), and the random-effects variance (v_θ), defined in equation (6). Taken together, this leads to the new variance, as shown in equation (7).

$$v_\theta = \frac{Q - (k - 1)}{\sum w_i - (\sum w_i^2 / \sum w_i)} \quad (6)$$

where the random-effects variance is defined as:

$$v_i^* = v_i + v_\theta \quad (7)$$

The new value of the variance, v_i^* , gives us a new standard error and a new inverse variance of every effect size. Re-estimating equations (1) and (2), using the new inverse variance, gives us a new weighted mean effect size.

A second way of solving the homogeneity problem is using a mixed-effects model, which combines both the fixed- and the random-effects methods. The mixed-effects method assumes that the subject-level sampling error is partly due to systematic factors within the total group of studies, and that it is partly due to the random sources. In other words, heterogeneity can partially be explained by differences in the observed characteristics of different studies. The mixed-effects method divides the studies into subgroups with a certain characteristic and calculates the weighted mean effect size per group. The mixed-effects method tests for homogeneity within each separate group, calculates the between-group homogeneity statistic, and lastly calculates the within-group homogeneity statistic per characteristic. If these homogeneity statistics are significant, the random-effects estimate per group is calculated.

Publication Bias

Publication bias is a type of selection bias, which can arise because not all research that has been done with respect to the subject at hand is published

or reported. The reasons for this could be that researchers and publishers do not trust the results or do not value the results properly. This particular type of bias is very likely to influence the results of a meta-analysis. Publication bias in a meta-analysis means that there are studies that cannot be found in common databases or books, which are 'missing' in the dataset of studies created for the meta-analysis. Due to these 'missing' studies, the weighted mean effect size might be biased. In our case, it is very likely that the weighted mean effect size is upward biased, as almost all values of the effect size are significant, and around or just above zero.

Tests for publication bias investigate whether the effect size is related to the sample size or the standard error of the estimate. If studies with a high t -value or a low p -value are more likely to be published, the bias will decrease with the standard error or sample size. Begg's test is an adjusted-rank correlation test, to check whether the effect size and its variance are correlated. Egger's test is based on a regression of the effect size on its inverse variance, and tests whether the intercept is significantly different from zero. Due to the parametric structure, it is more powerful than Begg's test. Doucouliagos and Stanley (2009) introduced the Funnel graph's Asymmetry Test (FAT) in which the t -value is regressed on the inverse variance. The estimate of the constant in this regression provides a test of publication bias. In absence of publication bias, the constant is to equal 0.

Publication bias can be corrected by estimating the effect size, controlling for the correlation between estimated effect size and the standard error of the estimate. In a nonparametric setting, this correction can be based on the so-called 'trim and fill' method (Duval and Tweedie 2000). This method copes with publication bias by trimming extreme positive studies and by adding some studies to the study database that appear to be missing, to make the dataset symmetric. After that, the weighted mean effect size is calculated again. A limitation of this method is that it assumes that the most extreme estimated effect sizes are the missing studies. Steichen (2000) studied the method and reports a slight tendency to overcorrection. The regression that is used to detect the relationship between the estimated effect size and its standard error, in a parametric setting, can also be used to estimate the average effect size where the standard error is 0, and thus publication bias can be assumed to be absent. This approach has been developed by Doucouliagos and Stanley (2009), and can easily be extended to a meta-regression context. Therefore, we discuss this approach in the next section.

Meta-regression Analysis

In a meta-regression, all the explanatory variables that are created during the coding stage can be regressed on the main variable, in this case, the estimated effect size. A standard meta-regression is a random-effects regression in which the relation between the background and the study characteristics and the effect size can be tested quantitatively. One of the advantages of a

meta-regression is that it has a multivariate context (Doucouliagos and Laroche 2003). A meta-regression makes it possible to investigate the relation between the effect size and one or more of the key variables, where the other variables, the study characteristics, can be used as controls.

However, the standard meta-regression has one major drawback: namely ignoring the existence of publication bias. If the meta-analysis is positively tested for publication bias, the meta-regression should also incorporate this bias. Because it is likely that the chance of being published is largely influenced by the significance level of the results of a study, an extra aspect has to be added to the meta-regression to account for this chance of publication. There are two possible methods to correct a meta-analysis for publication bias: (a) Hedges' (1992) maximum likelihood approach; and (b) the FAT-precision effect test (PET) method of Doucouliagos and Stanley (2009). Hedges (1992) developed an approach to estimate a meta-regression taking into account the effect of publication bias. This approach has been used, for example, in Ashenfelter *et al.* (1999). Hedges' meta-regression, which includes publication bias, assumes that there is a weight function, based on significance levels, which determines the probability that the study is observed. Studies with a high-significance level are more likely to be observed. To make it possible to derive the likelihood function, Hedges assumes that the probability that a study is published follows a step function. Assuming that each effect size is a random draw from a normal distribution with the true effect size as its mean and the documented standard error as its standard error, and assuming that the probability of being in the sample depends on the above-mentioned function of the t -value, Hedges' method is able to derive the likelihood function. The effect size itself is assumed to be a linear function of study characteristics plus a normally distributed error term. We refer to Hedges' article for further details on this topic. The basic idea is that there is a population of studies, each with a measured effect size that equals the true effect size for the group, plus an error term. This is represented in equation (8):

$$\widehat{ES}_i = ES_i + \eta_i \quad (8)$$

The true effect size depends on the characteristics of the training and the estimation technique used:

$$ES_i = P(\text{training characteristics, data, estimation method, etc.}) \quad (9)$$

$$t = \frac{\widehat{ES}}{\sigma_\eta} \quad (10)$$

The relative probability for each study to be rejected equals p_i . We define this function as a step function on five intervals:

$$\begin{aligned}
 g1, & \quad 2.054 \leq t < 2.576 & \Rightarrow p < 0.01; \\
 g2, & \quad 1.960 \leq t < 2.054 & \Rightarrow 0.01 \leq p < 0.05; \\
 p_i = g3, & \quad 1.645 \leq t < 1.960 & \Rightarrow 0.05 \leq p < 0.1; \\
 g4, & \quad 1.282 \leq t < 1.645 & \Rightarrow 0.1 \leq p < 0.2; \\
 g5, & \quad 0 \leq t < 1.282 & \Rightarrow 0.2 \leq p.
 \end{aligned}
 \tag{11}$$

We chose to use these five groups because these are the confidence levels that are mostly used. We furthermore allow for a *t*-value smaller than 0. The results are robust with respect to changes in these intervals.

Doucoulgiagos and Stanley (2009) use the FAT publication bias test as a basis for a meta-regression. In this test, publication bias is assumed to be a linear function of the standard error of the study.

The starting point is the following equation:

$$ES_i = \beta_1 + \beta_0 Se_i + \varepsilon_i
 \tag{12}$$

where *ES* is the effect size, *Se* is the standard error of the effect size and ε_i is the random error.

The random errors are assumed to be heteroskedastic, and *Se_i* is considered a measure of this heteroskedasticity. Therefore, equation (12) has to be divided by *Se_i*, which leads to:

$$t_i = \beta_0 + \beta_1(1/Se_i) + v_i
 \tag{13}$$

Doucoulgiagos and Stanley (2009) argue that a *t*-test of the intercept β_0 is a test for publication selection and may be considered the FAT. Furthermore, they argue that the coefficient of precision, β_1 , is the ‘true effect’ corrected for publication bias and the PET is used to test its significance. The FAT-PET method includes both the test for, and the correction of, publication bias.

The approach is equivalent to estimating:

$$ES_i = \beta_0 + \beta_1 Se_i + v_i
 \tag{14}$$

with weighted least squares using $(Se_i)^2$ as the weight. In equation (14), the intercept has to be interpreted as the ‘true’ effect size and the coefficient of precision β_1 as the indication for publication bias. Adding study characteristics to this equation turns this approach into a meta-regression that shows the link between the study characteristics and the effect size. Recently, Stanley and Doucoulgiagos (2007) propose to regress the effect size on the square of the standard error, again using $(Se_i)^2$ as the weight. This provides — based on their Monte Carlo experiments — better results than equation (14).

Both Hedges’ and Doucoulgiagos and Stanley’s method assume some structure in the relationship between the standard error and the bias. To investigate the performance of both methods, we carried out several Monte Carlo studies, varying the distribution of sample sizes and the function that

described the probability of studies to be selected. Our main findings are that, under most conditions, Hedges' method has no or only a very small bias, while the bias in Doucouliagos and Stanley's approach can still be substantial. Furthermore, the standard errors of Hedges' approach turn out to be smaller than standard errors using Doucouliagos and Stanley's approach. Regressing on the squared standard error (Stanley and Doucouliagos 2007) does indeed provide better results than regressing on the standard error of a study, but the performance of Hedges' method remains better. Also, the coefficients of the effects of study characteristics on the effect size have a smaller bias and lower standard error using Hedges' approach.

The reason for this difference in performance is that Doucouliagos and Stanley — in their weighting procedure — do not take into account variation caused by heterogeneity in effect sizes and, more important, the assumed functional form which states that the bias is linear in the standard error. However, Hedges' method also has its weaknesses. The assumed step function can be a poor representation of the true underlying publication function. The results can be particularly sensitive for improper choices of the cut-off points in this function. When there is no study within a certain group (g_1 – g_5), the method fails to converge. However, this is not the case in our analysis.

4. Meta-analysis results

Table 1 shows all the studies included in the meta-analysis, and Table 2 shows the weighted mean effect for all included studies. As can be seen in Table 2, the fixed-effects weighted mean effect size for all studies is 0.7 per cent, the unweighted mean effect size is 8.4 per cent and the random-effects weighted mean effect size is 3.9 per cent. The large difference between the unweighted mean effect size and the other effect sizes is mainly due to publication bias. There are several reasons to prefer the random-effects model over the other two. First, in the fixed-effects model, we have a very significant Q -statistic, implying heterogeneity in the effect sizes. Furthermore, the fixed-effects weighted mean effect size is for a large part calculated based on only a few studies. This suggests a possible bias towards the larger studies in the sample. One way of dealing with this possible bias is weighting all the studies equal when calculating the weighted mean effect size. However, this also means that insignificant studies and studies with a very low number of observations, or a low number of participants in the treatment or control group, are weighted the same as the very significant studies with more than 10,000 observations. Therefore, it is very unlikely that this approach would lead to a valuable estimate that we can rely on. Furthermore, we find an unrealistically high mean effect size of 8.4 per cent. Therefore, we turn to the random-effects model. The random-effects model gives us a much smoother distribution of weights and a parameter that is as significant as in the fixed-effects model. It is also a better method to use, considering the highly significant Q -statistic.

TABLE 1
Sources for Meta-Analysis

<i>Author and year</i>	<i>Number of effect sizes</i>	<i>First year of data</i>
Addison and Belfield (2008)	2	2004
Almeida-Santos and Mumford (2006)	1	1991
Almeida-Santos and Mumford (2010)	1	1991
Arriagada (1990)	2	1985
Arulampalam and Booth (2001)	2	1981
Bartel (1992)	2	1986
Bassanini and Brunello (2008)	1	1995
Beyer de (1990)	2	1980
Biesebroeck van (2007)	3	1991
Blau and Robins (1985)	2	1980
Booth (1991)	2	1987
Booth and Bryan (2005)	1	1998
Booth <i>et al.</i> (2003)	2	1991
Budria and Pereira (2004)	4	1998
Cohen-Goldner and Eckstein (2010)	2	1992
Evertsson (2004)	1	1994
Fredland and Little (1981)	1	1966
Gibson (2003)	2	1996
Gielen (2007)	1	1998
Groot (1995)	1	1952
Hamil-Luker (2005)	2	1979
Hill (1995)	2	1967
Kahyarara and Teal (2008)	2	1999
Kawaguchi (2006)	1	1994
Krueger and Rouse (1998)	2	1991
Kuckulenz and Zwick (2003)	1	1998
Leuven and Oosterbeek (2004)	4	1994
Leuven and Oosterbeek (2008)	2	2001
Marcotte (2000)	2	1966
Melero (2004)	2	1991
Metcalfe and Sloane (2007)	1	2004
Rosholm <i>et al.</i> (2007)	4	1995
Salas-Velasco (2009)	2	1994
Schöne (2004)	4	1989
Tan <i>et al.</i> (2007)	1	2005
Vignoles <i>et al.</i> (2004)	2	1991
Xiao (2002)	1	1996
Yoshida and Smith (2005)	1	1999
Total	71	—

Sensitivity Analysis Results

To test for the sensitivity of the meta-analysis results, we perform several sensitivity tests. These tests are done by excluding some of the studies. First, to attempt to exclude all outliers, we exclude the highest 5 per cent and 10 per cent and the lowest 5 per cent and 10 per cent of the studies with respect to both the effect size and the standard error of the effect size. Performing these analyses with the adapted datasets has hardly any effect on both the fixed and the random weighted mean effect size. We then exclude the top and bottom 5 and 10 per cent of the studies with respect to sample size. This analysis also

TABLE 2
(Weighted) Mean Effect Sizes

	<i>All studies</i>		
Number of studies	71	—	—
Range of sample size	168–34,900	—	—
Fixed-effects effect size	0.007	(0.001)	***
Unweighted effect size	0.084	(0.016)	***
Random-effects effect size	0.039	(0.003)	***
Heterogeneity random effects (<i>Q</i> -statistic)	631.71	***	—

Note: Standard errors in parentheses.

* Significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level.

TABLE 3
Egger's Test for Publication Bias

	<i>Coefficient</i>	<i>Standard error</i>	<i>t-stat</i>	<i>p > t </i>
Slope	0.002	0.001	1.500	0.139
Bias (intercept)	2.425	0.282	8.600	0.000
<i>N</i>	71	—	—	—

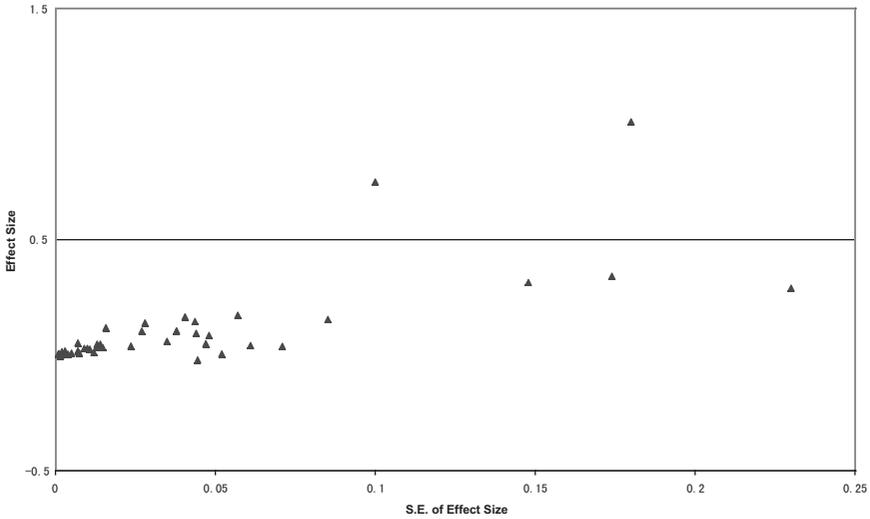
has a negligible effect on the fixed and random weighted mean effect size. The fixed-effects weighted mean effect size varies between 0.6 per cent and 1.2 per cent for all the analyses, whereas the random-effects weighted mean effect size varies between 3.9 per cent and 5.3 per cent. We find slightly larger changes in the case of excluding the top and bottom 10 per cent of the standard error of the effect size than for the other groups. However, the changes are still very small, and we see no reason to exclude any study from the meta-analysis that we have not excluded already in the study-selection stage.

Publication Bias Results

To check for publication bias, we use Egger's test and the FAT method of Doucouliagos and Stanley (2009). Egger's test has an intercept that is significantly different from zero and therefore also presents a strong indication of publication bias. The results of Egger's test can be found in Table 3. The plot of the effect size versus the standard error of the effect size in Figure 2 also shows the existence of publication bias. Furthermore, the simple FAT regression also points towards publication bias. We therefore conclude that publication bias exists in our dataset.

To correct for publication bias, we use what is called the 'trim and fill' method. The trim and fill method, when using the random-effects model, gives us a significant fixed-effects estimate of 0.6 per cent. The new estimate for the random-effects model has become 1.5 per cent. The trim and fill random-effects method added 22 studies to our original data set. We

FIGURE 2
Effect Size versus Standard Error (SE) of the Effect Size.



conclude that, taking into account the existence of publication bias, we have a difference in effects of training on wages between the trained and the non-trained of about 1.5 per cent. However, the trim and fill method has some drawbacks, and tends to over-correct, as described above, which is why we continue to study the wage effect of training by other methods that also take into account publication bias.

Separate Meta-analysis including Publication Bias

An alternative to the random-effects model is the mixed-effects model, for which the random-effects results for all subgroups can be found in Table 4. Because we have concluded that the random-effects model is to be preferred over the fixed-effects model, we only included the random-effects estimates in Table 4. We furthermore endorse the existence of publication bias, which is why we also correct for this bias (by the trim and fill method) in the second column of Table 4.

Table 4 shows the estimated weighted mean effect size, its standard error, and the corresponding significance level. Table 4 indicates that IV and FE studies have a smaller weighted mean effect size than OLS studies and studies that use other methods. It can be expected that the weighted mean effect size for IV and FE studies is lower than the weighted mean effect size of OLS studies. However, at the same time, the power of IV and FE estimates will also be lower than the power of OLS estimates. As a consequence, conditional on effect size, IV studies have a lower probability of being published than OLS studies. On the other hand, IV and FE are better methods, which

TABLE 4
Random-Effects Effect Sizes per Subgroup

	<i>Number of studies</i>	<i>Random-effect size</i>	<i>Standard error</i>		<i>Random-effect size corrected for publication bias</i>	<i>Standard error</i>	<i>—</i>
Total	71	0.038	(0.003)	***	0.015	(0.003)	***
Study characteristics							
Continent							
Europe	37	0.032	(0.003)	***	0.014	(0.004)	***
USA	13	0.061	(0.014)	***	0.011	(0.014)	**
Other	21	0.075	(0.015)	**	0.025	(0.015)	**
Way of precision							
SE	45	0.027	(0.003)	***	0.015	(0.003)	***
<i>t</i> -value	26	0.071	(0.009)	***	0.017	(0.009)	***
Estimation method							
IV/FE	21	0.025	(0.004)	**	0.012	(0.005)	**
OLS	41	0.044	(0.005)	***	0.014	(0.005)	***
Other	9	0.131	(0.025)	**	0.061	(0.024)	**
Total <i>N</i>							
<2000	35	0.061	(0.009)	***	0.015	(0.006)	***
>2000	36	0.038	(0.003)	***	0.012	(0.004)	***
Education control							
Included	59	0.048	(0.004)	***	0.014	(0.004)	***
Not included	12	0.019	(0.005)	***	0.013	(0.005)	***
Experience control							
Included	36	0.035	(0.004)	***	0.014	(0.001)	***
Not included	35	0.053	(0.006)	***	0.017	(0.006)	***
Tenure control							
Included	35	0.036	(0.004)	***	0.015	(0.004)	***
Not included	36	0.049	(0.005)	***	0.014	(0.006)	***
Article status							
Published	50	0.033	(0.004)	***	0.017	(0.004)	***
Not published	21	(0.052)	(0.006)	***	0.014	(0.006)	***
Sample characteristics							
Age							
15–65	55	0.038	(0.003)	***	0.014	(0.004)	***
<35	8	0.072	(0.021)	**	0.010	(0.020)	**
>35	8	0.049	(0.009)	***	0.047	(0.009)	***
Gender							
All	43	0.036	(0.003)	***	0.014	(0.004)	***
Male	17	0.050	(0.009)	***	0.018	(0.010)	**
Female	11	0.054	(0.013)	**	0.026	(0.014)	**
Wage							
Yearly, monthly and other wage	41	0.074	(0.008)	***	0.028	(0.008)	***
Hourly	30	0.019	(0.003)	***	0.010	(0.003)	***

Note: Standard errors (SE) in parentheses.

* Significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level. IV, instrumental variable; FE, fixed effect; OLS, ordinary least square.

may have a higher chance of getting published. The latter is also what we see in Table 4, that is, the difference between the corrected and the uncorrected IV and FE estimate is much larger than for OLS and other methods. We also see that using yearly or monthly wage instead hourly wage for the effect of on-the-job training on the growth of wages gives a higher weighted mean effect size. The difference gets smaller once corrected for publication bias.

Furthermore, studies with a smaller sample have a higher weighted mean effect size, as well as studies using *t*-statistics compared with standard errors as a way of precision. This might be due to differences in the publication biases for both types of studies. Indeed, the difference between standard errors and *t*-statistics gets smaller after correcting for publication bias. Another result is that women have a smaller effect size than men in both the model with and the model without correction for publication bias. We also

see that the effect for men and women separately is much larger than for the groups taken together. This may point at highly beneficial specific training programmes aiming at specific types of jobs.

Finally, US studies have a higher weighted mean effect size compared with European studies, and studies from other countries have an even higher effect. According to our results, separating the age groups gives higher results for the weighted mean effect size than using the complete working age group. We find that younger people have a higher weighted mean effect size than older people, and the complete age group. Once corrected for publication bias, we see not much difference between studies that include education, experience or tenure controls, and studies that do not include them.

5. Meta-regression analysis results

As explained before, the standard meta-regression analysis does not account for possible publication bias. Because we have detected publication bias in our study, we use the meta-regression with the maximum likelihood function as described in Hedges (1992) and equation (11) to account for selection bias. Furthermore, we also use the FAT-PET method with the squared standard error to correct for publication bias and compare both methods. The results of the meta-regression analyses of both methods can be found in Table 5.

In the FAT-PET models, the publication bias is corrected for by the parameter of the squared standard error of the effect size. Although previous tests report publication bias, this coefficient is insignificant in column 3. In Hedges' analysis, we can reveal the presence of publication bias by examining the parameters g_1 – g_5 . To test for this presence, we test the restriction of $g_1 = g_2 = g_3 = g_4 = g_5 = 1$ using a likelihood ratio test (column 2). Rejection of the test means that there is publication bias present in our model. This number exceeds the 1 per cent critical value of the chi-square test for all of Hedges' models presented in Table 5, implying that the hypothesis of no publication bias can be rejected. In Hedges' regression models as presented in Table 5, we see a clear difference in the size of g_1 – g_5 , in the corrected models (columns 6 and 8). We see that both g_4 and g_5 are very large in all cases, g_5 being larger than g_4 . We also see that g_2 and g_3 have a smaller effect size. Furthermore, only g_3 , g_4 and g_5 are significant, suggesting that studies with p -values between 0.01 and 0.05 have the same probability of being published as those with lower p -values. This result is very much like the result in the article of Ashenfelter *et al.* (1999).

Part 1 of Table 5 shows a comparison between Hedges' model and the FAT-PET model. A few aspects of this part of Table 5 are striking. First, note that the true effect size of the FAT-PAT model without correction for publication bias is equal to the fixed-effects effect size as presented in Table 2. Furthermore, note that the true effect size of the FAT-PET model and the Hedges' model, both corrected for publication bias, are *not* exactly the same. This difference is due to the weighting method used in Hedges' maximum

TABLE 5
Extended Regression Model Part 1

	1		2		3		4					
	Effect size	Standard error										
True effect size	0.007	(0.002)	***	0.056	(0.007)	***	0.007	(0.002)	***	0.026	(0.013)	**
σ	—	—	—	0.043	(0.006)	***	—	—	—	0.045	(0.006)	***
g_1	—	—	—	—	—	—	—	—	—	0	—	—
g_2	—	—	—	—	—	—	—	—	—	-0.231	(0.513)	—
g_3	—	—	—	—	—	—	—	—	—	0.469	(0.327)	—
g_4	—	—	—	—	—	—	—	—	—	0.672	(0.197)	***
g_5	—	—	—	—	—	—	—	—	—	0.833	(0.097)	***
Standard error of effect size squared	—	—	—	—	—	—	4.245	(2.612)	—	—	—	—
<i>N</i>	71	—	—	71	—	—	71	—	—	71	—	—
Gender	—	—	—	—	—	—	—	—	—	—	—	—
Reference = all	—	—	—	—	—	—	—	—	—	—	—	—
Male	—	—	—	—	—	—	—	—	—	—	—	—
Female	—	—	—	—	—	—	—	—	—	—	—	—
Wage	—	—	—	—	—	—	—	—	—	—	—	—
Reference = yearly, monthly and other	—	—	—	—	—	—	—	—	—	—	—	—
Hourly	—	—	—	—	—	—	—	—	—	—	—	—
Reference = IV/FE	—	—	—	—	—	—	—	—	—	—	—	—
Estimation method	—	—	—	—	—	—	—	—	—	—	—	—
OLS	—	—	—	—	—	—	—	—	—	—	—	—
Other	—	—	—	—	—	—	—	—	—	—	—	—
Way precision	—	—	—	—	—	—	—	—	—	—	—	—
Reference = SE	—	—	—	—	—	—	—	—	—	—	—	—
Means and <i>t</i> -value	—	—	—	—	—	—	—	—	—	—	—	—
Age	—	—	—	—	—	—	—	—	—	—	—	—
Reference = 15–65	—	—	—	—	—	—	—	—	—	—	—	—
<35	—	—	—	—	—	—	—	—	—	—	—	—
>35	—	—	—	—	—	—	—	—	—	—	—	—
Article status	—	—	—	—	—	—	—	—	—	—	—	—
Reference = not published	—	—	—	—	—	—	—	—	—	—	—	—
Published	—	—	—	—	—	—	—	—	—	—	—	—
First year of data	—	—	—	—	—	—	—	—	—	—	—	—
data used	—	—	—	—	—	—	—	—	—	—	—	—

Note: Standard errors (SE) in parentheses.

* Significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level. IV, instrumental variable; FE, fixed effect; OLS, ordinary least square; FAT-PET, Precision Effect Test-Funnel graph's Asymmetry Test.

TABLE 5
Extended Regression Model Part 2

	5		6		7		8	
	FAT-PET not corrected for publication bias		Hedges' not corrected for publication bias		FAT-PET		Hedges' corrected for publication bias	
	Effect size	Standard error	Effect size	Standard error	Effect size	Standard error	Effect size	Standard error
True effect size	0.018	(0.012)	0.072	(0.016)	0.019	(0.011)	0.050	(0.019)
σ	—	—	0.033	(0.004)	—	—	0.033	(0.005)
g_1	—	—	—	—	—	—	0	—
g_2	—	—	—	—	—	—	-0.046	(0.436)
g_3	—	—	—	—	—	—	0.558	(0.271)
g_4	—	—	—	—	—	—	0.729	(0.162)
g_5	—	—	—	—	—	—	0.865	(0.079)
Standard error of effect size squared	—	—	—	—	3.363	(2.338)	—	—
<i>N</i>	71		71		71		71	
Gender								
Reference = all	-0.009	(0.011)	-0.015	(0.015)	-0.009	(0.011)	-0.012	(0.017)
Male	0.002	(0.011)	-0.007	(0.016)	0.002	(0.011)	-0.006	(0.018)
Female								
Reference = yearly, monthly and other								
Hourly	-0.018	(0.010)	* -0.036	(0.012)	-0.017	(0.010)	* -0.034	(0.013)
Reference = IV/FE								
OLS	0.009	(0.008)	-0.015	(0.014)	0.009	(0.008)	-0.011	(0.015)
Other	0.055	(0.019)	*** 0.046	(0.020)	*** 0.054	(0.019)	** 0.052	(0.023)
Way precision								
Reference = SE								
Means and <i>t</i> -value	0.010	(0.011)	0.016	(0.012)	0.011	(0.011)	0.022	(0.014)
Age								
Reference = 15-65	0.012	(0.020)	0.020	(0.020)	0.012	(0.202)	0.034	(0.024)
<35	0.041	(0.026)	-0.006	(0.022)	0.039	(0.026)	-0.008	(0.023)
>35								
Article status								
Reference = not published								
Published	-0.001	(0.004)	-0.008	(0.013)	-0.002	(0.004)	-0.022	(0.015)
First year of data used	0.000	(0.000)	0.000	(0.001)	0.000	(0.000)	0.000	(0.001)

Note: Standard errors (SE) in parentheses.
 * Significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level.
 IV, instrumental variable; FE, fixed effect; OLS, ordinary least square; FAT-PET, Precision Effect Test-Funnel graph's Asymmetry Test.

likelihood estimation, which is not exactly the same as weighting by the standard error of the effect size squared, as is used in the FAT-PET model.

To estimate the models presented in part 2 of Table 5, we basically include all the variables that are shown in Table 4. These are the variables that intuitively may have an influence on the value of the effect size. Because the presence of the correction for publication bias (g_1 – g_5 and Se_i) already corrects for the p -value, we decided to leave the variable *Total N* out of the meta-regression. Furthermore, we leave the variable *continent* out of the regression. This is because this variable on its own does have an effect on the effect size, although insignificant, but in correlation with other variables, the effect of both *continent* and the other variables disappears. We see in Table 5 that some of the coefficients of the variables, for both methods, are not significantly different from zero, which could be explained by the relatively small N . However, we do believe that the size of the coefficients is realistic and definitely relevant.

As discussed in the methodological section, in general, Hedges' model performs better in most cases than the FAT-PET method, so we prefer Hedges' model. This means that the estimated wage returns to on-the-job training are 2.6 per cent, as can be concluded from regression 4 in part 1 of Table 5. For a matter of comparison and consistency with part 1 of Table 5, we also present both the FAT-PET (with squared standard errors) and the Hedges' model in part 2 of Table 5.

The most important regressions in Table 5 are regressions 7 and 8. These regressions show the results for the variables that are included in the model. As we have argued that we prefer Hedges' model over the FAT-PET method, we only discuss the results from meta-regressions 6 and 8. However, the results for the effects of the variables included in the model are similar. First, we see that the estimation methods IV and FE give a lower effect size than OLS and other methods. It seems that the better methods to determine the effect size result in lower effects. We furthermore see that the difference between OLS and other methods compared with IV and FE becomes smaller when we include the correction for publication bias. It seems that the power of the method used is very important to get satisfactory results.

Furthermore, Table 5 shows a significantly lower coefficient for the studies that have determined the effect of training on hourly wage, compared with the studies that have used monthly, yearly, or another type of wage as the dependent variable. Therefore, the quality of the data also seems to be an important characteristic.

Although not significant, we see that the separate groups of men and women give a lower effect size than for the two groups taken together, and that published articles have a lower effect size than unpublished articles. The latter difference becomes larger when corrected for publication bias. Furthermore, we find that age has no significant effect on the effect size. However, we see that the age group below 35 has a higher effect size than the other age groups. Then, Table 5 again shows that the studies that have used means and t -values for their estimation have a larger effect size than studies that have

used means and standard errors. This difference increases when correcting for publication bias. As the meta-regression assumes that the decision to publish only depends on the p -value, this may mean that studies using standard errors as way of precision to calculate the estimate have a higher chance of being published. Finally, we see that earlier studies have a slightly higher effect size, but this result is not significant either.

6. Concluding remarks

In this article, we have presented a meta-analysis of the wage effects of on-the-job training. There are four major problems with respect to investigating the wage returns to on-the-job training. First, most studies researching wage effects of training courses have a low power, which makes it statistically hard to obtain a reliable estimate. Second, due to this low power, studies on the wage effects of on-the-job training are disproportionately subject to publication bias. Third, not all studies take into account the selectivity issues with respect to who gets to participate in a training course, and who does not. Fourth, studies on the wage effects of on-the-job training are very diverse and have different characteristics, which, at first sight, make them hard to compare.

We used a meta-analysis to overcome these problems as a meta-analysis increases the sample size, takes into account publication bias, and makes it possible to compare the seemingly incomparable studies. The main finding of the study is that the average wage effect of on-the-job training is 2.6 per cent. We also find a substantial heterogeneity in the wage effects of training courses. Lastly, our findings show that an appropriate statistical methodology and high-quality data are needed to access these wage returns.

As said, this study finds an average wage effect of 2.6 per cent for on-the-job training. Comparing the average number of hours spent on on-the-job training with the average number of hours spent on schooling gives a wage increase of 30 per cent for on-the-job training, compared with 8 per cent for the return to schooling. Solely based on this aspect of this study, it might seem that on-the-job training is more worthwhile than investment in regular schooling. A potential source of this difference could be that people who invest in formal training in general also spend more time on informal learning (Borghans *et al.* 2011). After taking a course, people might spend some time at their work, experimenting with their newly acquired knowledge. The total time invested in training might therefore exceed the time measured in surveys. Furthermore, assuming an IRR of 8 per cent, the wage increase of 30 per cent indicates that, on average, on-the-job training is profitable up until the age of 61 if workers also have to pay for the investment. However, the return turns out to be substantially lower in studies that use estimation techniques that correct for selectivity bias. In that case, the maximum age at which a training course is profitable may reduce to 55.

We also find a substantial heterogeneity in the wage effects of training courses. This means that some courses are very profitable, maybe even at a

higher age than 61, whereas other courses are not profitable at all, not even at a very young age. The implication of this heterogeneity is that knowledge about the expected wage returns of a specific training course is important for a firm to decide about investments in on-the-job training.

Lastly, we find that the methodology and the quality of the data used for an analysis play a major role in determining the return to on-the-job training. The use of OLS versus IV, for example, makes a large difference in the results, as does the use of panel data. As several other studies have already emphasized, this implies that the results are very sensitive to the choice of models and methodology. Furthermore, the quality of the data plays a major role in determining the wage return to on-the-job training. Both the sample size and the selection of the participants largely determine the reliability of the results.

All in all, we can conclude that research and results on the wage return to regular schooling are far more stable than the returns to on-the-job-training. This might be due to, for example, the distinction between general and specific training, heterogeneity in who pays for the training, reasons for investments, and the duration and amount of training, all of which are often not taken into account in studies on the returns to training. The lack of this information on training leads to a larger study heterogeneity. Therefore, we should nuance the finding that on-the-job training seems to be more profitable than regular schooling until this can be proved based on more research. This meta-analysis has been carried out using relatively few studies of the effects of on-the-job training because these studies are scarce. More solid research on the wage effects of training would therefore be very useful. Our study has shown that the use of methods to correct for selectivity and the correction for publication bias is important, so this should explicitly be taken into account. Furthermore, to better compare studies on the wage effects of training, more precise information on the conditions of the training, the division of costs between employer and workers, and the nature of the training could help to improve our understanding of the determinants of these wage effects.

Final version accepted on 20 October 2011.

Acknowledgements

The authors would like to thank Jan Sauermann, Xander Koolman, Pierre Koning and two anonymous referees for their useful comments on previous versions of this article.

Notes

1. In 1998, a new tax law was implemented in the Netherlands, which introduced a tax deduction for a firms' expenditures on work-related training. Leuven and

- Oosterbeek (2004) used the tax deduction as instrumental variable to investigate training participation and the effect of training participation on wages.
2. We deliberately chose not to include a continuous measure for training for two reasons: first, there is only a small number of articles that present specific information on the duration of training. Second, studies that do present the duration of training often mention the training in weeks or months, and fail to specify the number of days in a month of training and the total hours per week of training. As we cannot compare these studies, as they use different standards, we decided not to include continuous measures of training at all.
 3. As the *t*-statistic shows the significance at first glance, whereas the standard error requires a closer look, it could be possible that this may also cause publication bias.

References

- Addison, J. T. and Belfield, C. R. (2008). 'Unions, training and firm performance'. IZA Discussion Papers 3294, Bonn: Institute for the Study of Labor.
- Almeida, R. and Carneiro, P. (2006). 'Costs, Benefits and the Internal Rate of Return to Firm Provided Training'. The World Bank Policy Research Working Paper Series.
- Almeida-Santos, F. and Mumford, K. (2006). 'Employee training, wage dispersion and equality in Britain'. IZA Discussion Papers 2276, Bonn: Institute for the Study of Labor.
- and — (2010). 'Employee training and wage dispersion: white and blue collar workers in Britain'. IZA Discussion Papers 4821, Bonn: Institute for the Study of Labor.
- Arriagada, A.-M. (1990). 'Labor market outcomes of non-formal training for male and female workers in Peru'. *Economics of Education Review*, 9 (4): 331–42.
- Arulampalam, W. and Booth, A. L. (2001). 'Learning and earnings: Do multiple training events pay? A decade of evidence from a cohort of young British men'. *Econometrica*, 68 (271): 379–400.
- Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999). 'A review of estimates of the schooling/earnings relationship, with test for publication bias'. *Labour Economics*, 6 (4): 453–70.
- Barret, A. and O'Connell, P. J. (2001). 'Does training generally work? The returns to in-company training'. *Industrial and Labor Relations Review*, 54 (3): 647–62.
- Barron, J. M., Berger, M. C. and Black, D. A. (1998). 'Do workers pay for on-the-job training?' *Journal of Human Resources*, XXXIV (2): 235–52.
- Bartel, A. P. (1992). 'Training, wage growth and job performance: evidence from a company database'. NBER Working Paper series. Cambridge, MA: National Bureau of Economic Research.
- Bassanini, A. and Brunello, G. (2008). 'Is training more frequent when the wage premium is smaller? Evidence from the European Community Household Panel'. *Labour Economics*, 15: 272–90.
- Becker, G. (1964). *Human Capital*. New York: Columbia University Press.
- Ben-Porath, Y. (1967). 'The production of human capital and the life-cycle of earnings'. *Journal of Political Economy*, 75 (4): 352–65.
- Beyer de, J. (1990). 'The incidence and impact on earnings of formal training provided by enterprises in Kenya and Tanzania'. *Economics of Education Review*, 9 (4): 321–30.

- Biesebroek van, J. (2007). Wage and productivity premiums in sub-Saharan Africa. NBER Working Paper series. Cambridge, MA: National Bureau of Economic Research.
- Blau, D. M. and Robins, P. K. (1985). 'Communications, training programs and wages. A general equilibrium analysis of the effects of program size'. *The Journal of Human Resources*, XXII (1): 113–25.
- Booth, A. L. (1991). 'Job-related formal training: who receives it and what is it worth?' *Oxford Bulletin of Economics and Statistics*, 53 (3): 281–94.
- and Bryan, M. L. (2005). 'Testing some predictions of human capital theory: new training evidence from Britain'. *The Review of Economics and Statistics*, 87 (2): 391–4.
- , Francesconi, M. and Zoega, G. (2003). 'Unions, work-related training, and wages: evidence for British men'. *Industrial and Labor Relations Review*, 57 (1): 68–91.
- Borghans, L., Golsteyn, B. H. H. and de Grip, A. (2011). 'Measuring Learning'. Mimeo, Maastricht University.
- Budria, S. and Pereira, P. T. (2004). 'On the returns to training in Portugal'. IZA Discussion Papers 1429, Bonn: Institute for the Study of Labor.
- Cohen-Goldner, S. and Eckstein, Z. (2010). 'Estimating the return to training and occupational experience: the case of female immigrants'. *Journal of Econometrics*, 156: 86–105.
- Conti, G. (2005). 'Training, productivity and wages in Italy'. *Labour Economics*, 12: 557–76.
- Dearden, L., Reed, H. and van Reenen, J. (2006). 'The impact of training on productivity and wages: evidence from British panel data'. *Oxford Bulletin of Economics and Statistics*, 67 (4): 397–421.
- Doucouliaqos, C. and Laroche, P. (2003). 'What do unions do to productivity? A meta-analysis'. *Industrial Relations*, 42 (4): 650–91.
- and Stanley, T. D. (2009). 'Publication selection bias in minimum-wage research? A meta-regression analysis'. *British Journal of Industrial Relations*, 47 (2): 406–28.
- Duval, S. and Tweedie, R. (2000). 'Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis'. *Biometrics*, 56: 455–63.
- Evertsson, M. (2004). Formal on-the-job training: a gender-typed experience and wage-related advantage? *European Sociological Review*, 20 (1): 79–94.
- Frazis, H. and Loewenstein, M. (2006). 'Wage Compression and Divisions of Returns to Productivity Growth: Evidence Form EOPP'. Bureau of Labor Statistics Working Paper, Bureau of Labour Statistics.
- Fredland, J. E. and Little, R. D. (1981). 'Self-employed workers: returns to education and training'. *Economics of Education Review*, 1 (3): 315–37.
- Gibson, J. (2003). 'Do lower expected wage benefits explain ethnic gaps in job-related training? Evidence from New Zealand'. Motu Working Paper 2003-03.
- Gielen, A. C. (2007). 'Performance pay, training and labor mobility'. IZA Discussion Papers 2932, Bonn: Institute for the Study of Labor.
- Greenberg, D. H., Michalopoulos, C. and Robins, P. K. (2003). 'A meta-analysis of government-sponsored training'. *Industrial and Labor Relations Review*, 57 (1): 31–53.
- , — and — (2006). 'Do experimental and non experimental evaluations give different answers about the effectiveness of government-funded training programs?' *Journal of Policy Analysis and Management*, 25 (3): 523–52.

- Griliches, Z. (1977). 'Estimating the returns to schooling: some econometric problems'. *Econometrica*, 45 (1): 1–22.
- Groot, W. (1995). 'The wage effects of investments in enterprise-related-training'. *Empirical Economics*, 20: 133–47.
- (1999). 'Productivity effects of enterprise-related training'. *Applied Economics Letters*, 6: 369–71.
- Hamil-Luker, J. (2005). 'Women's wages: cohort differences in returns to education and training over time'. *Social Science Quarterly*, 86 (Supplement): 1261–78.
- Heckman, J. J., LaLonde, R. J. and Smith, J. A. (1999). 'The economics and econometrics of active labor market programs'. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*. 3A. New York: North-Holland, pp. 1865–2097.
- Hedges, L. V. (1992). 'Modeling publication selection effects in meta-analysis'. *Statistical Science*, 7 (2): 246–56.
- and Olskin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Hill, E. T. (1995). 'Labor market effects of women's post-school-age training'. *Industrial and Labor Relations Review*, 49 (1): 138–49.
- Kahyarara, G. and Teal, F. (2008). 'The returns to vocational training and academic education: evidence from Tanzania'. *World Development*, 36 (11): 2223–42.
- Kawaguchi, D. (2006). 'The incidence and effect of job training among Japanese women'. *Industrial Relations*, 45 (3): 469–77.
- Krueger, A. and Rouse, C. (1998). 'The effect of workplace education on earnings, turnover and job performance'. *Journal of Labor Economics*, 16 (1): 61–94.
- Kuckulenz, A. and Zwick, T. (2003). 'The Impact of Training on Earnings-Differences between Participant Group and Training Forms'. ZEW Discussion Paper.
- Leuven, E. and Oosterbeek, H. (2004). 'Evaluating the effect of tax deductions on training'. *Journal of Labor Economics*, 22 (2): 461–88.
- and — (2008). 'An alternative approach to estimate the wage returns to private-sector training'. *Journal of Applied Econometrics*, 4: 423–34.
- Loewenstein, M. and Spletzer, J. (1997). 'Delayed formal on-the-job training'. *Industrial and Labor Relations Review*, 51 (1): 82–99.
- and — (1999). 'General and specific training: evidence and implications'. *Journal of Human Resources*, 34 (4): 710–33.
- Marcotte, D. E. (2000). 'Continuing education job training and the growth of earnings inequality'. *Industrial and Labor Relations Review*, 53 (4): 602–23.
- Melero, E. (2004). 'Evidence on Training and Career Paths: Human Capital, Information and Incentives'. IZA Discussion Papers 1377, Bonn: Institute for the Study of Labor.
- Metcalf, R. and Sloane, P. J. (2007). Human capital spillovers and economic performance in the workplace in 2004: some British evidence. IZA Discussion Papers 2774, Bonn: Institute for the Study of Labor.
- Mincer, J. (1958). 'Investment in human capital and personal income distribution'. *Journal of Political Economy*, 66 (4): 281–302.
- OECD (2000). *Final Report of the International Adult Literacy Survey*. Paris: OECD.
- (2003). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- Rosholm, M., Skyt Nielsen, H. and Dabalén, A. (2007). 'Evaluation of training in African enterprises'. *Journal of Development Economics*, 84 (1): 310–29.

- Salas-Velasco, M. (2009). 'Beyond lectures and tutorials: formal on-the-job training received by young European university graduates'. *Research in Economics*, 63: 200–11.
- Schøne, P. (2004). 'Firm-financed training: firm specific or general skills?' *Empirical Economics*, 29: 885–900.
- Schonewille, M. (2001). 'Does training generally work? — Explaining labour productivity effects from schooling and training'. *International Journal of Manpower*, 22 (1/2): 158–72.
- Stanley, T. D. and Doucouliagos, C. (2007). 'Identifying and Correcting Publication Selection Bias in the Efficiency-Wage Literature: Heckman Meta-regression'. School Working Paper, No. 2007/11, School of Accounting, Economics and Finance, Victoria, Australia.
- Statistics Netherlands (2000). *Time-Budget Survey 2000 — TBO 2000 (2000)*. Amsterdam: Steinmetz Archives.
- Steichen, T. J. (2000). 'sbe39: nonparametric trim and fill analysis of publication bias in meta-analysis'. *Stata Technical Bulletin*, STB57: 8–14.
- Tan, H., Savchenko, Y., Gimpelson, V., Kapelyushnikov, R. and Lukyanova, A. (2007). 'Skills shortage and training in Russian enterprises'. IZA Discussion Papers 2751, Bonn: Institute for the Study of Labor.
- Vignoles, A., Galindo-Rueda, F. and Feinstein, L. (2004). 'The labour market impact of adult education and training: a cohort analysis'. *Scottish Journal of Political Economy*, 51 (2): 266–80.
- Xiao, J. (2002). 'Determinants of salary growth in Shenzhen, China: an analysis of formal education, on-the-job training and adult education with a three-level model'. *Economics of Education Review*, 21: 557–77.
- Yoshida, Y. and Smith, M. R. (2005). 'Training and earnings of immigrant males: evidence from the Canadian workplace and employee survey'. *Social Science Quarterly*, 86 (Supplement): 1218–41.