

---

Is Training more frequent when the Wage Premium is  
smaller? Evidence from the European Community

Andrea Bassanini and Giorgio Brunello

---

Unclassified

DELSA/ELSA/WD/SEM(2006)10



Organisation de Coopération et de Développement Economiques  
Organisation for Economic Co-operation and Development

05-Dec-2006

English text only

**DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS  
EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS COMMITTEE**

DELSA/ELSA/WD/SEM(2006)10  
Unclassified

**OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS No. 41**

**IS TRAINING MORE FREQUENT WHEN THE WAGE PREMIUM IS SMALLER?  
EVIDENCE FROM THE EUROPEAN COMMUNITY HOUSEHOLD PANEL**

**Andrea Bassanini and Giorgio Brunello**

*JEL Classification: J24, J31, J41*

All Social, Employment and Migration Working Papers are now available through OECD's Internet website at <http://www.oecd.org/els>

**JT03219255**

Document complet disponible sur OLIS dans son format d'origine  
Complete document available on OLIS in its original format

English text only

**DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS**

<http://www.oecd.org/els>

**OECD SOCIAL, EMPLOYMENT AND MIGRATION  
WORKING PAPERS**

<http://www.oecd.org/els/workingpapers>

This series is designed to make available to a wider readership selected labour market, social policy and migration studies prepared for use within the OECD. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language – English or French – with a summary in the other.

Comment on the series is welcome, and should be sent to the Directorate for Employment, Labour and Social Affairs, 2, rue André-Pascal, 75775 PARIS CEDEX 16, France.

The opinions expressed and arguments employed here are the responsibility of the author(s) and do not necessarily reflect those of the OECD

**Applications for permission to reproduce or translate  
all or part of this material should be made to:**

**Head of Publications Service  
OECD  
2, rue André-Pascal  
75775 Paris, CEDEX 16  
France**

**Copyright OECD 2006**

## ACKNOWLEDGEMENTS

Andrea Bassanini is Economist at the Directorate for Employment, Labour and Social Affairs of the OECD and Associate Professor at the Department of Economics (CEPN) of the University of Paris XIII. Giorgio Brunello is Professor at the Department of Economics of the University of Padova and Research Fellow at IZA, Bonn and CESifo, Munich. Comments and suggestions by Daron Acemoglu, Giuseppe Bertola, Federico Biagi, Alison Booth, Michael Burda, Eve Caroli, Daniele Checchi, Maria De Paola, Martine Durand, John Martin, Raffaele Miniaci, Hessel Oosterbeek, Enrico Rettore, Jean-Marc Robin, Muriel Roger, Paolo Sestito, Ken Swinnerton, Alain Trannoy, Tanguy van Ypersele, Guglielmo Weber, Lüdger Wössmann, two anonymous referees as well as seminar participants in Aix, Evry, Lyon, Marseille, Milan, Munich, Orléans, Novara, Padova, Paris, Pisa and Tokyo are gratefully acknowledged. The views expressed in this paper are those of the authors and do not necessarily reflect those of the OECD or its member countries.

## SUMMARY

According to Becker [1964], when labour markets are perfectly competitive, general training is paid by the worker, who reaps all the benefits from the investment. Therefore, *ceteris paribus*, the greater the training wage premium, the greater the investment in general training. Using data from the European Community Household Panel, we compute a proxy of the training wage premium in clusters of homogeneous workers and find that smaller premia induce greater incidence of off-site training, which is likely to impart general skills. Our findings suggest that the Becker model provides insufficient guidance to understand empirical training patterns. Conversely, they are not inconsistent with theories of training in imperfectly competitive labour markets, in which firms may be willing to finance general training if the wage structure is compressed, that is, if the increase in productivity after training is greater than the increase in pay.

## RESUME

Dans la théorie de Becker [1964], lorsque le marché du travail est en concurrence parfaite, seuls les salariés investissent dans la formation générale, car ils sont les seuls à pouvoir s'approprier les retombées bénéfiques de la formation. Par conséquent, toutes choses égales par ailleurs, plus la prime salariale à la formation est élevée et plus l'investissement en formation est importante. Sur la base des données du Panel Communautaire des Ménages, nous calculons une proxy de la prime salariale à la formation pour des groupes homogènes de salariés et nous trouvons une relation inverse entre cette proxy et l'incidence de la formation hors site, qui concerne, selon toute vraisemblance, des compétences relativement générales. Nos résultats suggèrent que le modèle de Becker ne fournit pas une clé interprétative suffisante pour comprendre les tendances empiriques de la formation. Par contre, la distribution de la formation ne semble pas être en contradiction avec les théories de la formation qui prévoient que, lorsque le marché du travail est en concurrence imparfaite, les entreprises peuvent être disposées à investir en formation si l'augmentation de la productivité qui en découle est supérieure à l'augmentation du salaire.

**TABLE OF CONTENTS**

SUMMARY .....4

RESUME.....5

IS TRAINING MORE FREQUENT WHEN THE WAGE PREMIUM IS SMALLER? EVIDENCE FROM  
THE EUROPEAN COMMUNITY HOUSEHOLD PANEL.....7

1. Introduction.....7

2. General Training and Wage Premia.....9

3. Data .....13

4. The Empirical Results .....17

5. Conclusions.....25

APPENDIX .....27

REFERENCES .....37

## IS TRAINING MORE FREQUENT WHEN THE WAGE PREMIUM IS SMALLER? EVIDENCE FROM THE EUROPEAN COMMUNITY HOUSEHOLD PANEL

### 1. Introduction

1. Human capital is a key determinant of economic growth. The amount of training individuals receive during their working life has a significant impact on their career prospects, wages and employability. Moreover, improving workers' competencies is crucial in the face of rapid technological change. In spite of the broad consensus on the importance of training, there is a large debate in the economic literature and in policy circles concerning whether the current level of investment in training is efficient and which agent (employer or employee) has the greater incentive to invest.

2. According to Becker [1964], when labour markets are perfectly competitive, *general* training — that is training which raises productivity at other employers to the same extent as at the employer who provides it — is fully paid by the worker, who reaps all the benefits from the investment. However, although few surveys have information on the generality of skills, in those that do, most of the reported job-related training appears to be employer-paid, at least partially, even when it is viewed by respondents as general (Barron, Berger and Black [1999], Loewenstein and Spletzer [1999], Booth and Bryan [2005]). Furthermore, in many surveys, employers appear to pay for most of off-site training (OECD [2003]), which is found by Loewenstein and Spletzer [1999] to be essentially general. This evidence is difficult to reconcile with Becker's model unless we argue that the employee fully compensates the employer by accepting lower wages during (or before) training spells. Yet, no clear empirical support exists for this fact in the literature.<sup>1</sup>

3. Recent theories of imperfect competition in the labour market can explain why employers have an incentive to pay for apparently general human capital. If the market for trained workers is less competitive than the market for untrained workers, the ratio of wages to productivity is lower for trained than for untrained workers — that is, wages are compressed with respect to productivity along the training dimension. In these circumstances, the employer has an incentive to train because he can afford to pay a trained worker less than the marginal product while still retaining her (see *e.g.* Katz and Zidermann [1990], Stevens [1994], Acemoglu and Pischke [1999a], Lazear [2003] and Booth and Zoega [2004]).

4. In this paper, we contribute to a growing body of literature attempting at discriminating empirically between these two alternative theories of training, by estimating how the probability of receiving general training is affected by the training wage premium — that we approximate with the difference between the median wage growth of trained and untrained employees in clusters of relatively homogeneous workers. Becker's model implies a positive relationship between the training wage premium and the incidence of general training. By contrast, a negative relationship falls within the possible outcomes of training theories based on imperfect competition in the labour market.

---

1. See among others, Loewenstein and Spletzer [1998], Barron, Berger and Black [1999] and Sicilian [2001], as well as Bishop [1997] for a survey of earlier studies.

5. A few recent papers have provided empirical evidence that is consistent with models of firm-sponsored training based on imperfect competition in the labour market. In particular, panel data studies, which control for individual fixed effects and for job mobility, show that some of the benefits of general training are appropriated by workers with some lag and/or when they change employers. This evidence is consistent with the view that employers have some monopsony power over their own trained workers. For example, using data from the US National Longitudinal Survey of Youth (NLSY), Loewenstein and Spletzer [1999] find that, when training imparts general skills, the estimated effect of completed spells of employer-paid training on earnings is three times larger for training spells completed during previous jobs than during the current job. Similarly, Loewenstein and Spletzer [1998] find that completed spells of employer-provided off-site training in the current job have no effect on current wages. By contrast, off-site employer-paid training received at previous employers has a positive and persistent impact on wages. Using more waves of the same data, Lengermann [1999] finds that the latter effect increases over time. Booth and Bryan [2005] study three recent waves of the British Household Panel Survey (BHPS) and find that employer-provided training has a positive and persistent impact on wages, with evidence that the impact is larger for accredited training received at previous employers. Similar results are obtained by Blundell, Dearden and Meghir [1999], using three distant waves of the British National Child Development Survey (NCDS), and by Gerfin [2004] on Swiss data.

6. However, the presence of labour market imperfections is not the only possible explanation of the finding that wages after a training spell grow faster if the worker changes jobs. For instance, workers might undertake training to qualify for jobs in other firms that are more efficient at employing trained workers (see *e.g.* Moen and Rosen [2002]), and the main empirical predictions of the Becker model are therefore not necessarily in contrast with these findings.

7. Another way of discriminating between alternative theories of training is to investigate the relationship between the minimum wage and training incidence. According to Becker, high minimum wages should decrease investments in human capital, as they would prevent minimum-wage workers from accepting wage cuts to finance training. By contrast, in the alternative theories of training, the greater the minimum wage, the greater is the incentive for firms to pay for general training. The reason is that, in an imperfectly competitive labour market, the minimum wage compresses the lower tail of the wage distribution without necessarily affecting individual productivity (see *e.g.* Acemoglu and Pischke [2003]). However, recent empirical studies in the United States and in the United Kingdom report contradictory findings on the impact of the minimum wage on training.<sup>2</sup> There are several possible reasons why this strand of research is inconclusive. For instance, in countries where the minimum wage is high, it might be difficult to find a group which is not directly or indirectly affected by the minimum wage and qualifies as a genuine control. Conversely, in countries where the minimum wage is particularly low, the incidence of training in the treatment group is likely to be extremely small, since training is relatively infrequent at the bottom of the wage distribution. Moreover, most of these studies focus on changes of the minimum wage over time, but it is not clear what time horizon is appropriate to analyze the effect of institutional changes such as increments in the minimum wage. Last but not least, the degree of imperfection of the labour market might differ across countries.

8. The approach we follow in this paper is similar to that taken by studies of minimum wage and training, insofar as we focus on the implications of the different theories on the relationship between wage premia and training. We use cross-country data from the European Community Household Panel (ECHP) and partition workers into clusters of relatively homogeneous employees (in terms of country, education, occupation and sector). We then construct cluster-specific measures of the training wage premium – that we compute as the difference between the median wage growth rates of trained and untrained employees –

---

2. See Grossberg and Sicilian [1999], Neumark and Wascher [2001], and Acemoglu and Pischke [2003], for the United States, and Arulampalam, Booth and Bryan [2004], for the United Kingdom.

and investigate whether these measures have a significant impact on general training, after dealing with the potential problems of endogeneity. We find a negative relationship between our measure of the training wage premium and the probability of taking general training. Our findings are inconsistent with the Becker model, which implies a positive relationship between the training wage premium and the incidence of general training. Conversely, although we cannot observe individual productivity gains from training, our findings can be interpreted as lower bound estimates of the impact of wage compression on training, to the extent that the (unobserved) productivity premium and the (observed) wage premium are not negatively correlated. Under this condition, our findings are consistent with theories based on imperfect competition. This approach seems to us particularly suitable to analyse the empirical relevance of different theories of training using European data: since the migration of labour between EU countries is still limited, we can meaningfully use the country dimension in the definition of clusters, which allows us to construct a sample with a large number of clusters and to obtain significant variation in our measures of the training premium.<sup>3</sup>

9. The paper is organized as follows. Section 2 presents our empirical approach, Section 3 describes the data and Section 4 is dedicated to the presentation of the empirical results. Conclusions follow.

## 2. General Training and Wage Premia

10. To fix the ideas and discuss our empirical specification, in this section we briefly sketch a simplified two-period model of general training, which essentially corresponds to the Becker model in the case of perfect competition, and to a simplified version of the model by Acemoglu and Pischke [1999a] in the case of imperfect competition.<sup>4</sup>

11. In the first period a worker is matched with an employer and training can take place. In the second period, the worker can quit if she receives a better wage offer. Let us denote with  $f(\tau)$ ,  $w(\tau)$  and  $v(\tau)$  the worker's second-period labour productivity, wage and outside option, respectively, with  $\tau = 1$  indicating that the individual takes general training in the first period (while  $\tau = 0$  indicates that no training takes place). Let us assume that training is perfectly general, which is equivalent to assuming that  $f(\tau)$  and  $v(\tau)$  do not vary across firms. In the case of perfect competition considered by Becker,  $f(\tau) = w(\tau) = v(\tau)$  and  $\Delta f = \Delta w = \Delta v$ , where  $\Delta w = w(1) - w(0)$  is the wage premium and  $\Delta f$  and  $\Delta v$  are similarly defined. Since  $\Delta f - \Delta w = 0$ , if training costs  $c(I)$  are positive and  $c(0)=0$  the employer does not invest. Conversely, and assuming no discounting for simplicity, the worker invests and pays the training costs as long as  $\Delta w > c(1)$ . If training costs vary across workers, the greater the wage premium, the larger is the number of employees that are ready to pay for training. In other words, for a given individual  $i$ , the probability that she decides to take training can be written as:

$$\Pr ob\{\tau_i = 1\} = \Pr ob\{\Delta w_i > c_i(1)\} \quad [1]$$

---

3. A testing approach based on estimating the impact of minimum wages on training can be difficult to implement when continental European countries are added to the sample. In countries such as Austria, Germany and Italy, there are no statutory minimum wages, but wage floors are defined by collective agreements, often at a very disaggregate sectoral level, and it is not easy to collect precise information on them.

4. See also Booth and Zoega [2004] for a very similar model. Most of the other models based on imperfect competition retain similar features.

12. If we remove the assumption of perfect competition, the worker has to incur a cost upon quitting — that is, her outside option  $v(\tau)$  is below her productivity  $f(\tau)$  by the amount  $D(\tau) > 0$ . Assume that the current employer wants to retain the trained worker. He can offer the worker a wage  $w$  equal to her outside option  $v$  plus a fraction of the gap  $D$ , that is  $w(1) = v(1) + \beta(f(1) - v(1)) = v(1) + \beta D(1)$ , where  $\beta < 1$  represents the worker's bargaining power. In the absence of training, the wage offered to retain her would have been  $w(0) = v(0) + \beta(f(0) - v(0)) = v(0) + \beta D(0)$ . Assuming that workers cannot bear any training cost (e.g. because of borrowing constraints), the expected profits from training and no training are  $\pi(1) = f(1) - w(1) - c(1) = (1 - \beta)(f(1) - v(1)) - c(1)$  and  $\pi(0) = f(0) - w(0) = (1 - \beta)(f(0) - v(0))$ , respectively. In the decision to invest in general training, the employer compares the expected profits from training with the expected profits in the event of no training. The employer decides to bear the training costs if and only if  $\Delta f - \Delta w > c(1)$  or, equivalently,  $\Delta f - \Delta v > c(1)/(1 - \beta)$ .<sup>5</sup> If training costs vary across workers, the more compressed is the wage structure (the higher  $\Delta f - \Delta w$ ), the larger the number of employees for which employers are ready to pay for training. For a given individual  $i$ , the probability that she receive employer-paid training is:

$$\Pr ob\{\tau_i = 1\} = \Pr ob\{\Delta f_i - \Delta w_i > c_i(1)\} \quad [2]$$

13. With a similar argument, it can be shown that, even when workers bear some of the training cost, the employer's investment in general training increases with wage compression. In this case, however, total investment in training may or may not increase with wage compression, since the incentives for the employee to invest in training are greater, the greater the training wage premium  $\Delta w$ .<sup>6</sup> Thus, the case of perfect competition considered by Becker can be viewed as a special case of the Acemoglu and Pischke model: in the absence of labour market frictions,  $D$  is equal to zero,  $f(\tau) = w(\tau) = v(\tau)$  and the employer never pays for general training. Taking into account that the training that is reported to be employer-sponsored might be indirectly paid by the worker by accepting lower wages during training, equations [1] and [2] can be generalised as:

$$\Pr ob\{\tau_i = 1\} = \Pr ob\{\gamma \Delta f_i + \sigma \Delta w_i + \delta c_i(1) > 0\} \quad [3]$$

with  $\gamma > 0$ ,  $\sigma \geq -\gamma$  and  $\delta < 0$ , in the case of the Acemoglu and Pischke model,<sup>7</sup> and  $\gamma = 0$ ,  $\sigma > 0$  and  $\delta < 0$ , in the case of the Becker model.

- 
5. There are several types of labour market imperfections that might make  $D$  an increasing function of training. See Acemoglu and Pischke [1999b] for a survey. Lazear [2003] suggests that this pattern might also emerge if skills are a multi-dimensional variable, with different firms using different combinations of general skills, except when many firms use the same combination of skills and workers can search the whole space of alternative jobs, which would correspond to the case of perfect competition.
  6. In this respect, Acemoglu and Pischke's results are similar to those by Hashimoto in his model of firm-specific training, where the parties share the costs and benefits of the investment (see Hashimoto [1981]).
  7. The parameter  $\sigma$  can be strictly higher than  $-\gamma$  if employer-provided training is partly financed by the employee.

14. Employers and employees decide on the investment in training by forming expectations about the training wage premium. We posit that agents form their expectations by looking at the current wage distribution for trained and untrained employees. The heterogeneity of workers and jobs suggests, however, that the relevant distribution should not be the entire wage distribution, but rather an appropriate portion of it. Therefore, we partition individuals into relatively homogeneous clusters and approximate the training wage premium with the difference between the median wage growth rates of those who reported to have received training in the period covered by the survey and of those who did not<sup>8</sup>. Individual wages are affected both by the current training investment and by the accumulated training stock before the sample period. Since we can observe individual training history only for very few years, a substantial part of this stock is not observed. By using *growth* rates rather than *levels*, we are able to eliminate the influence on wages of the training stock accumulated before the reference period. We call this measure the wage growth differential.

15. Without cross-country comparable matched employer-employee datasets (see Bartelsman, Scarpetta and Schivardi [2005]), there are no cross-country data on individual productivity. Therefore, we treat the productivity gain from training  $\Delta f$  as an omitted variable and discriminate among competing theories on the basis of the estimate of the parameter  $\sigma$ . This implies that if we find  $\hat{\sigma} < 0$  (with  $\hat{\sigma}$  standing for estimate), then we will conclude that our evidence is in contrast with the empirical prediction of the Becker model. Conversely, we will not be able to discriminate among different theories if we find  $\hat{\sigma} \geq 0$ .

16. How does the omission of the productivity gain  $\Delta f$  bias our estimates? If the Becker model were true, omitting the productivity gain from the specification would induce no bias, since its expected coefficient is zero. Conversely, in the Acemoglu and Pischke model, to the extent that productivity and wage premia are not negatively correlated, the omission of the former would at most bias our estimates of  $\sigma$  against finding a negative relationship between the wage premium  $\Delta w$  and training incidence. In this case, our estimate of  $\sigma$  could be interpreted as a lower bound estimate of the impact of wage compression on training.

17. We do not observe training costs directly but assume that they vary with firm size, individual age, tenure in the job, educational attainment, type of labour contract, previous unemployment record, sector of activity, occupation and country. We also include a set of variables related to the employee's household, such as the household structure, family responsibilities and the financial situation of the household (we will call these variables "family variables" hereafter), which are intended to capture the worker's opportunity costs of training and liquidity constraints. Letting  $X$  be the vector which includes all these controls, we shall assume that  $\delta c_i(1) = \beta X_i + \varepsilon_i$  in equation [3].<sup>9</sup>

---

8. The reason why we use medians instead of averages is to reduce the weight of outliers. Nevertheless, we have performed a sensitivity analysis using average wage growth rates and found qualitatively similar results (available from authors upon request).

9. We experimented also with some interaction terms and other controls in the sensitivity analysis.

18. Letting  $\Delta W_c$  be the wage growth differential in cluster  $c$ , our empirical specification becomes

$$\Pr ob\{\tau_{ic} = 1\} = \Pr ob\{\alpha + \beta X_{ic} + \sigma \Delta W_c + \varepsilon_{ic} > 0\} \quad [4]$$

where the null hypothesis is  $\sigma \geq 0$ , as implied by the Becker model.

19. A potential objection to our empirical proxy of the wage premium is that it can be endogenous. For instance, if there are diminishing returns to human capital and trained and untrained workers are imperfect substitutes in production, we expect that the greater the *stock* of previously accumulated training the lower the productivity gain from training. Therefore, to the extent that the *stock* of previously accumulated training and the *flow* of training taken in the period covered by our data are correlated, we may find a negative relationship between the training wage premium and the probability of training even if the Becker's model is true.<sup>10</sup>

20. To take this issue into account, we test the hypothesis of weak exogeneity of the wage growth differential by following the methodology suggested by Smith and Blundell [1986]. To implement this test, we need to select at least one exogenous variable which is correlated with  $\Delta W_c$  but is independent of training incidence (or the probability of training), conditional on  $\Delta W_c$ . Our selected instrument is the difference between the log median age of those who have received training and the log median age of those who have not (we call this difference the log age differential and denote it with  $\Delta A_c$ ).<sup>11</sup>

21. We expect  $\Delta A_c$  and  $\Delta W_c$  to be correlated because training wage premia have been found to decrease with age, at least in European countries (see *e.g.* OECD, 2004). This result is typically explained by the fact that, in the absence of training, the wedge between productivity and the wage tends to decrease with age (see *e.g.* Abowd and Kramarz, 2003 and Aubert and Crepon, 2004, for France), due to labour market institutions and sectoral characteristics that affect the degree of downward nominal wage rigidity as well as the extent of internal labour markets. As workers get older, thus, training simply allows matching non-decreasing wages with otherwise decreasing productivity, with no apparent wage premium.

22. We argue that  $\Delta A_c$  also satisfies the orthogonality condition for instrument validity. First, there is no compelling reason for a direct causal impact of  $\Delta A_c$  (which is an aggregate age variable) on the probability of training, once we have controlled for observable individual characteristics, including individual age.<sup>12</sup> Second,  $\Delta A_c$  is not affected by diminishing returns by construction: with diminishing returns, an increase in the number of trained employees is likely to affect the wage of all workers, but has no influence on demographic characteristics such as age. True, since younger workers are likely to be trained first, we expect that the greater the incidence of training, the greater is the median (log) age of workers who receive training as well as of those who do not. But the log age differential  $\Delta A_c$  will not be affected if the two distributions — of the log age of workers who receive training and of those who do not

---

10. Another potential source of endogeneity is measurement error. See Bassanini and Brunello [2003] for other examples.

11. This measure roughly corresponds to the percentage difference between the two medians divided by 100.

12. A direct effect might emerge in the presence of large human capital externalities, if the intensity of these externalities varies with age. We exclude this possibility since we are aware of no evidence supporting this hypothesis in the literature.

— have the same degree of concentration around their respective medians.<sup>13</sup> If this is the case — as we show it is at the end of the next section — diminishing returns affect relative prices  $\Delta W_c$  without affecting relative ages  $\Delta A_c$ .

### 3. Data

23. We use individual data from the December 2001 release of the *European Community Household Panel* (ECHP), which is a longitudinal survey modelled on the British Household Panel Survey (BHPS). This survey provides a wealth of information on individual income and socio-economic characteristics for all EU countries and aims to be representative both in cross-sections and longitudinally. Due to the common questionnaire, the information contained in the ECHP is, in principle, comparable across countries, which is its main strength.

24. We use training data from the 1996 wave of the ECHP for seven countries<sup>14</sup> and restrict our attention to male employees (excluding apprentices),<sup>15</sup> aged from 30 to 60 years<sup>16</sup> and working full-time in the non-agricultural private sector, excluding sectors where non-profit organizations account for a non-negligible fraction of employment.<sup>17</sup>

25. The main question on vocational training in the ECHP is as follows “Have you at any time since January (year before the survey year) been in any vocational education or training, including part-time and short-courses?”. From this question, we construct a dichotomous variable “training participation”, which

- 
13. To see this, assume first that, in a given cluster, an individual is trained if and only if her age is below a given threshold, and that increasing training incidence implies shifting this threshold upwards, thereby making the youngest individuals in the group of the untrained switch to the group of the trained, of which they become the oldest. The crucial point is that both the median log ages of trained and untrained workers are lifted up, but the size of their relative increase depends on the concentration of the log age distribution in the neighbourhood of the two medians. If the distribution has the same degree of concentration in the neighbourhood of the two medians, the difference between them is independent from either the size of the groups and/or the exact age of switching individuals, and therefore the aggregate training rate (this statement is derived formally in the appendix). The same argument applies if the age of the individuals switching groups because of a marginal increase in training incidence is below the median of the untrained and above that of the trained. Furthermore, the argument can be easily extended to the case in which the number of switching individuals whose age is above both medians is not systematically different from the number of switching individuals whose age is below both medians. This suffices to our purposes. In fact, if age negatively affects the probability of training in a significant way, we expect the age of the bulk of switching individuals to be comprised between the two medians, with no systematic departure in either one direction for the others. Conversely, if age does not affect the probability of training, there will be no problem of reverse causality potentially plaguing the validity of the instrument, by construction.
14. The choice of the survey year and the country sample is dictated by data availability (see appendix). Included countries are Austria, Belgium, France, Germany, Italy, Spain, and the United Kingdom. In the sensitivity analysis we explore the robustness of results to changes in the survey year.
15. We exclude female employees to dampen problems of selection into employment.
16. We exclude individuals under 30 to reduce the risk that our results be altered by different national institutions affecting the school-to-work transition (such as different apprenticeship systems, with different degrees of government support). We also perform a sensitivity analysis by varying the age of the sample.
17. This choice of sectors is due to the fact that we cannot distinguish between business enterprises and non-profit organizations. The model described in the previous section may not apply to the latter. Included sectors are mining, manufacturing, utilities, construction, wholesale and retail trade, hotels and restaurants, transport, storage and communication, financial services, real estates, renting and business activities (corresponding to NACE Rev. 1 codes C to K).

takes value 1 if the individual responded “yes” and 0 if she responded “no”. Conditional on a positive answer, the individual is asked to report additional information on the last course only (including the type, whether it is still ongoing at the date of the survey and whether it is paid for or provided by the employer). If more than one concurrent course is involved, only the information concerning the course considered by the respondent as the most important is reported. Table 1 shows training events and training incidence — trained individuals as a percentage of the relevant population — by country and selected characteristics. About 17% of the individuals in our sample have experienced training, but there is a large cross-country variation. The country ranking shown in the table is similar to that which emerges from other cross-country European surveys, such as Eurostat Continuing Vocational Training Survey (CVTS) (see Nestler and Kailis [2002] and OECD [2003]).

**Table 1. Training events in the sample, by country and selected characteristics.**

Country	Individuals receiving training (in % of the total)	Other Characteristics	Individual receiving training (in % of the total)
Austria	159 (21.1)	Less than upper sec. educati	201 (7.7)
Belgium	110 (19.9)	Upper secondary education	550 (18.4)
France	201 (15.5)	More than upper sec. educati	443 (33.7)
Germany	219 (19.9)	Mining, manuf. and utilities	653 (15.0)
Italy	81 (6.7)	Services	541 (21.4)
Spain	150 (11.8)	High-skilled occupations	622 (31.0)
United Kingdom	274 (39.1)	Medium-skilled occupations	538 (12.4)
Total	1194 (17.3)	Low-skilled occupations	34 (6.3)

Note: male employees, aged from 30 to 60 years and working full-time in the non-agricultural private sector, excluding sectors where non-profit organizations have a non-negligible share of employment. The table shows the number of individuals who reported to have received training in the 1996 survey, by country and selected characteristics. Training incidence (trained individuals as a percentage of the relevant population) is reported in parentheses.

26. Respondents who have been in vocational education or training are asked to select the type of training received among the following categories: *a)* third level qualification, such as technical college; *b)* specific vocational training at a vocational school or college; *c)* specific vocational training within a system providing both work experience and complementary instruction elsewhere; *d)* specific vocational training in a working environment, without complementary instruction elsewhere; *e)* other. The distribution of training events by type is 5.1% for type a, 16.3% for type b, 11.1% for type c, 64.6% for type d, and 2.8% for type e. The questionnaire also asks individuals whether the last vocational training course was paid for or organized by the employer. As expected, about 86.9% of the courses on which the information is available are paid for or organized by the employer. Interestingly, the more formal the training the lower the employer support: 57% for third level qualification, such as technical college; 76% for specific vocational training at a vocational school or college; 92% for specific vocational training within a system providing both work experience and complementary instruction elsewhere; and 95% for vocational training in a working environment. This appears to be consistent with the conjecture that on-the-job training tends to be less general than off-the-job training.

27. As said in the previous section, our empirical measure of the training premium is based on comparing the wage growth rates of individuals who reported to have received training in the period covered by the survey and of those who did not — hereafter we will refer to them as “trained” and “untrained”, respectively. Individual wage growth is computed as the difference between the log gross hourly wage reported in the current wave – year 1996 – and the log wage reported in the previous wave – year 1995 – by the same individual. In the case of Austria, for which 1995 wage data are not comparable

with those from other years (see the Appendix), we replace the 1995-1996 wage growth with that of 1996-1997. Additional information on wages is reported in the Appendix.

28. Table 2 shows the difference between the median wage growth rates of trained and untrained individuals by country, educational attainment, occupation, and sector of activity. The median wage growth of trained individuals is lower than the median wage growth of the untrained in Germany and Italy, although the difference in these countries is close to zero. The table shows a wage return in the range 0-2%. Although these figures might seem low, they are consistent with panel data estimates based on European data (see Pischke [2001], Booth and Bryan [2005], Gerfin [2004], Schøne [2004] and OECD [2004]). The simple comparison of Table 2 and Table 1 shows no clear correlation pattern between training and wage premia at this aggregate level.

**Table 2. Wage growth gaps between trained and untrained workers.**

Country	Median wage growth, difference trained / untrained (%)	Other Characteristics	Median wage growth, difference trained / untrained (%)
Austria	1.05 [1.32]	Less than upper sec. education	-0.08 [1.50]
Belgium	0.33 [1.85]	Upper secondary education	0.34 [0.35]
France	1.38 [0.72]	More than upper sec. education	1.17 [0.93]
Germany	-0.17 [0.97]	Mining, manuf. and utilities	1.15 [0.63]
Italy	-0.21 [2.68]	Services	0.00 [0.63]
Spain	4.52 [2.32]	High-skilled occupations	-0.24 [0.90]
United Kingdom	0.05 [1.48]	Medium-skilled occupations	0.61 [0.52]
Average	0.29 [0.23]	Low-skilled occupations	5.19 [5.63]

Note: Percentage-point difference between the median wage growth rates of those who reported to have received training in the period covered by the 1996 survey (1997 for Austria) and of those who did not. The sample is limited to full-time male employees aged from 30 to 60 years and working in the non-agricultural business sector, excluding individuals who reported to be still in training at the time of the survey. Wage data refer to 1995-1996 for all countries except for Austria, for which they refer to 1996-1997. Medians are weighted by cross-sectional weights. Bootstrapped standard errors, obtained with 100 replications, in brackets.

29. We compute our empirical proxy of the training wage premium (what we called the “wage growth differential” in the previous section and denoted with  $\Delta W_c$  in equation [4]) by cluster. We define clusters by four dimensions: the country, the educational attainment (*less than upper secondary, upper secondary, more than upper secondary*),<sup>18</sup> the broad group of sectors (*mining, manufacturing, utilities and construction, and services*),<sup>19</sup> and the broad occupational group (*high-skilled occupations and medium- and low-skilled occupations*).<sup>20</sup> By so doing we obtain 12 clusters per country, but some of them are empty.<sup>21</sup>

18. Corresponding to ISCED categories, 0-2, 3, and 5-7, respectively.

19. Corresponding to NACE Rev. 1 industries C to F, and G and above, respectively.

20. Corresponding to managers, professional technicians and associate professional — ISCO-88 codes 1 to 3 — and to clerks, service and sales workers, craft and related trade workers, plant and machine operators and assemblers, and elementary occupations — ISCO-88 codes 4 to 9, respectively. Low-skilled occupations are aggregated to medium-skilled occupations since the size and training incidence of this occupational group is too small to be used separately in the definition of clusters (see Table 1).

21. Given that in our data sectors and, especially, occupations are defined at a lower level of aggregation (see appendix), it might be argued that a finer partition of the data should be used to define clusters. However, since

30. The wage growth differential  $\Delta W_c$  is obtained by computing for each cluster the difference between the median wage growth rates of the employees who reported to have received training in the period covered by the survey<sup>22</sup> and of those who did not, excluding workers who either changed cluster between the two wage observations or have missing cluster affiliation in one of the relevant interviews (about 6.5% of the sample). Since our measure of the wage growth differential should capture the cluster-specific premium to completed training spells, we exclude from this calculation also all individuals who reported that they were still in training at the time of the survey or have missing information on this question (about 19.4% of individuals who received training in the period covered by the survey). Although we limit our regression analysis to workers employed in sectors where non-profit organizations are not important, we consider all employees in the non-public service sector in the computation of the wage growth differential. This is done because we believe that the whole private service sector should be considered as the relevant market for service sector workers.<sup>23</sup> Finally, in order to reduce the weight of outliers, we drop clusters with 30 observations or less and with less than five reported training events. After all these exclusions we are left with 47 valid clusters.<sup>24</sup> Overall, we observe no clear pattern of cross-cluster bivariate correlation between training incidence and the wage growth differential (the correlation coefficient between these variables at the cluster level being only -0.13).

31. One problem with this measure of the training wage premium is that, given the formulation of the questions on training in the ECHP questionnaire (see above), when we compute the wage growth differential we misclassify as trained certain individuals who in fact received training only between January of the year before the current survey and the date of the previous survey (that is, except for Austria, between January 1995 and the date of the 1995 survey). As shown by Frazis and Loewenstein [1999], this implies that the computed wage growth differential will tend to be smaller than the value we would obtain if we focused only on training taken between the two survey dates (that is the dates of the two wage observations). Notice, however, that, to the extent that there is no reason why the proportion of misclassified individuals in a cluster should vary with training incidence, we can expect measurement error not to be systematically related to training incidence at the cluster level. In other words, this type of measurement error will, at most, bias our results towards zero, thereby making a rejection of the Becker model more difficult. Moreover, the absence of a systematic relationship between training incidence and measurement error can be tested by following the same instrumental variable procedure suggested by Smith and Blundell [1986], that we discussed in the previous section.

32. In the previous section we argued that the difference in the log median age of trained and untrained employees (the log age differential  $\Delta A_c$ ) is a valid instrument for the wage growth differential  $\Delta W_c$ . We noted, however, that this statement is correct only if there is no systematic difference in the concentration of the log age distributions in the neighbourhood of the median age of those who received

---

our empirical measure of the training premium controls for time-invariant heterogeneity within clusters (see below), the error induced by heterogeneity in large clusters is likely to be smaller than the error induced by small cluster size. Nevertheless, in the sensitivity analysis, we explore the effect of different partitions by varying the grouping of occupations.

22. That is, for growth rates between 1995 and 1996, those who reported to have received training in the 1996 survey and, for growth rates between 1996 and 1997, those who reported to have received training in the 1997 survey.
23. We check, however, the robustness of our results by repeating the analysis using wage growth differentials computed only for those sectors that we use in the final regressions.
24. These threshold limits reduce the number of clusters by more than one third. Not surprisingly, however, retained clusters account for a much larger share of observations with non-missing wages (see appendix). Nevertheless, we check the robustness of our results by lifting these size thresholds.

training and of those who did not. These hypotheses can be verified by computing for each cluster, and for two quantiles that are not far from the median, the interquantile difference for both age distributions and by checking whether their cross-cluster averages are systematically different. We performed this test for the difference between the 55<sup>th</sup> and the 45<sup>th</sup> percentiles, the 60<sup>th</sup> and the 40<sup>th</sup> percentile and the 70<sup>th</sup> and the 30<sup>th</sup> percentiles. As shown in Table 3, no significant difference emerges for any of the three, which supports the validity of  $\Delta A_c$  as an instrument.

33. Further details on the construction of the variables used in the empirical analysis as well as descriptive statistics are reported in the Appendix.

**Table 3. Tests of differences in the concentration of the log age distribution around the medians of the trained and the untrained.**

Interquantile difference	Cross-cluster average difference
P55-P45	-0.007 (0.010)
P60-P40	-0.011 (0.013)
P70-P30	-0.012 (0.016)

Note: The table reports the cross-cluster average differences between any given interquantile difference for the log age distribution of those who received training and the same interquantile difference for the log age distribution of those who did not receive training. Averages are weighted by cluster size. Standard errors in parentheses

#### 4. The Empirical Results

34. Our empirical analysis proceeds in two steps. First, we estimate a probit model of the probability of taking any type of training, using the cluster - specific wage growth differential  $\Delta W_c$  as our measure of the training wage premium, and testing its weak exogeneity with respect to training. Second, we re-estimate our model by focusing exclusively on training that is likely to be general. A series of robustness checks on the latter specification are presented at the end of the section and in the Appendix.

35. Table 4 shows the results of the first model, in which the dependent variable, training participation  $T$ , is equal to one in the event of training and to zero in the event of no training. The table is divided into two panels: Panel A refers to all training with no distinction in terms of financing; in Panel B we also use the available information on who financed the last training course and repeat the analysis by focusing on the probability of receiving employer-sponsored training.<sup>25</sup> In this case, we redefine the dependent variable  $T$  and set it equal to one in the event of employer – provided training and to zero in the event of no training.

25. While we do not know whether workers indirectly paid for employer-sponsored training by accepting lower wages, we can be relatively confident that employers did not pay for courses that are reported to be non-sponsored. It might therefore be desirable to eliminate these events from our dependent variable, since alternative theories have identical predictions for them under all circumstances. When the last training course is not reported to be employer-sponsored, we set therefore  $T$  to missing since we do not know whether the individual has taken employer-sponsored courses before the last one in the period covered by the survey.

36. We take explicitly into account the fact that the training variable  $T$  and the wage growth differential  $\Delta W_c$  are measured at different levels of aggregation and adjust the standard errors by allowing errors to be independent between clusters and correlated within clusters (as suggested by Moulton [1986]). The table reports only the relevant coefficients from each equation. The estimated equation corresponding to Panel A is reported in the Appendix. Full results for all the other models are available from the authors on request.

37. We consider three specifications: the first specification includes only basic controls (family variables, age, education, occupation, sector and country; Columns 1, 4, 7 and 10), the second adds to these controls firm size (Columns 2, 5, 8 and 11), and the third includes an extended set of controls (adding tenure, permanent job and previous unemployment; Columns 3, 6, 9 and 12). The first three columns of each panel report standard estimates, while the other columns report IV estimates obtained with the procedure suggested by Smith and Blundell [1986]: we compute the residual from the first-stage regression in which the wage differential is regressed on the cluster-specific instrument  $\Delta A_c$  and on country, occupation, education and sector dummies; then we add this variable to the probit specification and re-estimate it. The estimated coefficient of  $\Delta W_c$  that we obtain with this procedure is a consistent IV estimate of the impact of the wage growth differential. Furthermore, we can also test the weak exogeneity of the wage growth differential  $\Delta W_c$  by checking whether the estimated coefficient associated to the residual from the first-stage regression is statistically different from zero.

38. The results that emerge from Table 4 are inconclusive. On the one hand, standard estimates point to no correlation between the wage growth differential  $\Delta W_c$  and the probability of training. On the other hand, IV estimates tend to suggest a much more negative and often significant relationship.<sup>26</sup> Exogeneity tests, however, are not clear-cut as regards which estimation procedure should be preferred. In fact, even though in Panel A residuals are insignificant, thereby not rejecting the exogeneity hypothesis, they are often large in absolute terms. Furthermore, according to the results reported in Panel B, the exogeneity hypothesis is rejected at the 10% level in two specifications.<sup>27</sup>

---

26. One possible explanation of this difference is that standard estimates are strongly affected by measurement error, while diminishing returns is a second-order problem.

27. The ambiguity of these results is not surprising given that our instrument is relatively weak: the value of the F-test on the significance of the instrument is 5.0, which is relatively low (although statistically significant at the 5% level).

Table 4. *Probits for total training*. Dependent variable: training participation *T* in 1996.

Panel A: all training						
	(1)	(2)	(3)	(4)	(5)	(6)
Wage growth differential	-0.193 (0.411)	-0.168 (0.464)	-0.345 (0.417)	-1.890 (1.295)	-2.318* (1.319)	-2.386** (1.219)
Residual from first-stage regression				1.777 (1.393)	2.254 (1.450)	2.148 (1.344)
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Controls for firm size	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>no</i>	<i>yes</i>	<i>yes</i>
Extended controls	<i>no</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>no</i>	<i>yes</i>
Number of observations	5898	5728	5515	5898	5728	5515
F-test on instrument significance	5.00	5.00	5.00	5.00	5.00	5.00
Pseudo R-squared	0.154	0.171	0.179	0.154	0.171	0.179
Panel B: Employer-sponsored training						
	(7)	(8)	(9)	(10)	(11)	(12)
Wage growth differential	0.009 (0.509)	0.040 (0.574)	-0.216 (0.513)	-1.940 (1.388)	-2.545* (1.410)	-2.537** (1.288)
Residual from first-stage regression				2.036 (1.520)	2.706* (1.593)	2.436* (1.456)
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Controls for firm size	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>no</i>	<i>yes</i>	<i>yes</i>
Extended controls	<i>no</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>no</i>	<i>yes</i>
Number of observations	5791	5623	5412	5791	5623	5412
F-test on instrument significance	5.00	5.00	5.00	5.00	5.00	5.00
Pseudo R-squared	0.167	0.190	0.201	0.168	0.191	0.202

Note: Cluster-adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Individuals reporting non-sponsored training are excluded from Panel B. Each column of each panel refers to a different specification. Basic controls are country, education, occupation and sector dummies plus age and family variables. Specifications with extended controls add tenure, permanent job status and previous unemployment dummies to firm size dummies and basic controls. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. \*\*\*, \*\*, \*: significant at the 1%, 5% and 10% level of confidence, respectively.

39. In Table 4 we do not distinguish between general and firm-specific training. This is unsatisfactory for two reasons. First, alternative theories have different predictions on the relationship between the training wage premium and training only with regard to general training. Second, the inconclusive results reported in the table might be driven by the heterogeneity of training.<sup>28</sup>

40. We do not have in our dataset information on the generality of skills imparted through training. However, as noted in the previous section, respondents who have been in vocational education or training are asked to classify their last course in one (and only one) of the following mutually exclusive categories: *a)* third-level qualification, such as technical college; *b)* specific vocational training at a vocational school or college; *c)* specific vocational training within a system providing both work experience and complementary instruction elsewhere; *d)* specific vocational training in a working environment, without complementary instruction elsewhere; *e)* other. Following Loewenstein and Spletzer [1998, 1999] and OECD [2003] we can use the distinction between off-site and workplace training to proxy the distinction between general and firm-specific training. Assuming that all off-site training is at least partly general, we treat categories a, b and c as “general training” and category d as “firm-specific training”.<sup>29</sup> Nevertheless, since training falling under category c is partly taken off-site and partly received in the workplace, we also experiment with a different classification by assigning category c to “firm-specific training”. Throughout the remainder of the paper, we will refer to the former classification as “extensive definition of general training” and to the latter as “restrictive definition of general training”.

41. Table 5 reports results obtained by re-estimating the models of Table 4 (Panel A) using general training as the dependent variable. In this case, we exclude the observations where the last training course is reported to be firm – specific, since we do not know whether these individuals also received general training in the period covered by the survey (as discussed above, only one answer is allowed in the questionnaire)<sup>30</sup>. Three results stand out. First, under both definitions, standard estimates suggest a statistically significant negative correlation between the wage growth differential  $\Delta W_c$  and the probability of receiving general training.<sup>31</sup> Second, since the coefficient of the residual from the first stage is relatively small and statistically insignificant, we cannot reject the null hypothesis of weak exogeneity of the wage growth differential  $\Delta W_c$ . Third, once the residual is added to the specification, the estimated coefficient of the wage growth differential becomes less statistically significant. Notice, however, that the inclusion of the residual leads to inefficient estimates under the hypothesis of weak exogeneity.<sup>32</sup> Overall, these results

---

28. For instance, if the true wage premium to firm-specific training is close to zero or at least much smaller than the true wage premium to general training, our wage growth differential  $\Delta W_c$  will tend to be more correlated with the latter than with the former. Therefore, the greater the share of firm-specific training in total training in our sample, the greater is the importance of measurement error.

29. The option “other” cannot be classified and we choose to drop it from the sample.

30. The key results in Table 5 are not affected in a qualitative way if we explicitly account for both general and firm-specific training and use multinomial logit models. See the Appendix for details.

31. One might argue that we should use as wage growth differential the difference between the wage growth of those who received general training and of those who received no training. However, due to the relatively low incidence of general training, we would be left with only 15 valid clusters after applying our size thresholds. For this reason, we do so only in the sensitivity analysis reported in the Appendix, with no qualitative change of results.

32. This argument appears more compelling in the case of the extensive definition, where the residual from the first stage is close to zero. However, when the restrictive definition is used, IV estimates remain significant at least at the 10% level. In interpreting this figure, note that it might be argued that given the structure of our

suggest that there is a negative relationship between the training wage premium and the individual probability of attending a general training course. These findings appear inconsistent with the Becker model, which predicts a positive relationship between the training wage premium and the incidence of general training. The estimated impact, however, is rather small: taking, for example, the estimates in the first panel, we obtain that doubling the wage growth differential from its sample mean (0.195) would reduce the probability of general training (using its “extensive” definition) by only 3.1%.

42. In Table 6 and Chart 1, we check the robustness of our results to several modifications of the base sample (additional sensitivity analysis — as regards specification, estimation method, computation of the wage growth differential and further variations in the sample — is presented in the Appendix). In Panels A and B of Table 6, we report results obtained using employer-sponsored general training as the dependent variable (excluding employees who reported that the last course taken was non-sponsored); in Panel C we lift the restriction of five trained individuals per cluster; and in Panels D and E we modify the age range of the sample by including employees aged from 20 to 30 years. Overall, the results presented in the table are consistent with those of Table 5: standard estimates of the coefficient of the wage growth differential are always significant and relatively stable across specifications/samples; residuals from the first-stage regression are often insignificant (except when the sample is composed of employees aged between 20 and 50 years); whenever the residual is significant and/or large in absolute terms, the IV estimate of the coefficient of the wage growth differential is always negative and significant.

43. The small number of countries (7) in our sample suggests the possibility that our results might be driven by a single country. We assess the robustness of our findings to variations in country coverage by eliminating one country at a time and re-estimating our models. Chart 1 reports the results of this sensitivity analysis for the estimated coefficients of the impact of the wage growth differential on the propensity to take general training, using the extensive definition of general training (with extended controls only). Coefficient estimates are substantially robust to variation of country coverage.

---

null hypothesis a one-tail test would be more appropriate than a two-tail test. P-values derived on the basis of two-tail tests can be seen as lower bounds to the actual P-value.

Table 5. *Probits for general training*. Dependent variable: Participation in general training in 1996.

Panel A: Extensive definition						
	(1)	(2)	(3)	(4)	(5)	(6)
Wage growth differential	-1.719*** (0.561)	-1.757*** (0.574)	-1.821*** (0.542)	-1.995 (1.566)	-2.342 (1.652)	-2.229 (1.552)
Residual from first-stage regression				0.284 (1.579)	0.602 (1.672)	0.422 (1.590)
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Controls for firm size	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>no</i>	<i>yes</i>	<i>yes</i>
Extended controls	<i>no</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>no</i>	<i>yes</i>
Number of observations	5178	5028	4857	5178	5028	4857
F-test on instrument significance	5.00	5.00	5.00	5.00	5.00	5.00
Pseudo R-squared	0.226	0.230	0.243	0.226	0.230	0.243
Panel B: Restrictive definition						
	(7)	(8)	(9)	(10)	(11)	(12)
Wage growth differential	-1.900** (0.775)	-1.907** (0.774)	-2.018** (0.787)	-3.585* (1.853)	-3.671* (1.891)	-3.561* (1.891)
Residual from first-stage regression				1.741 (1.649)	1.821 (1.695)	1.592 (1.681)
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Controls for firm size	<i>no</i>	<i>yes</i>	<i>yes</i>	<i>no</i>	<i>yes</i>	<i>yes</i>
Extended controls	<i>no</i>	<i>no</i>	<i>yes</i>	<i>no</i>	<i>no</i>	<i>yes</i>
Number of observations	5090	4940	4771	5090	4940	4771
F-test on instrument significance	5.00	5.00	5.00	5.00	5.00	5.00
Pseudo R-squared	0.244	0.243	0.250	0.244	0.244	0.250

Note: Cluster-adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Individuals reporting firm-specific training are excluded from the sample. Each column of each panel refers to a different specification. The extensive definition of general training is used in Panel A, the restrictive one in Panel B. Basic controls are country, education, occupation and sector dummies plus age and family variables. Specifications with extended controls add tenure, permanent job status and previous unemployment dummies to firm size dummies and basic controls. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. \*\*\*, \*\*, \*: significant at the 1%, 5% and 10% level of confidence, respectively.

**Table 6. Robustness checks. Dependent variable: Participation in general training in 1996.**

Panel A: Employer-sponsored training, extensive definition, base sample						
	(1)	(2)	(3)	(4)	(5)	(6)
Wage growth differential	-1.658**	-1.784**	-1.922**	-3.228*	-3.952**	-3.608*
	(0.776)	(0.796)	(0.754)	(1.876)	(1.937)	(1.891)
Residual from first-stage regression				1.615	2.224	1.739
				(1.696)	(1.732)	(1.719)
Basic controls	yes	yes	yes	yes	yes	yes
Controls for firm size	no	yes	yes	no	yes	yes
Extended controls	no	no	yes	no	no	yes
Number of observations	4131	4103	4076	4131	4103	4076
F-test on instrument significance	7.14	7.14	7.14	7.14	7.14	7.14
Pseudo R-squared	0.271	0.290	0.304	0.271	0.291	0.304
Panel B: Employer-sponsored training, restrictive definition, base sample						
	(7)	(8)	(9)	(10)	(11)	(12)
Wage growth differential	-1.672*	-1.771*	-1.951**	-4.181*	-4.539*	-4.008*
	(0.935)	(0.961)	(0.923)	(2.376)	(2.473)	(2.474)
Residual from first-stage regression				2.582	2.841	2.103
				(2.144)	(2.242)	(2.249)
Basic controls	yes	yes	yes	yes	yes	yes
Controls for firm size	no	yes	yes	no	yes	yes
Extended controls	no	no	yes	no	no	yes
Number of observations	3920	3895	3871	3920	3895	3871
F-test on instrument significance	7.14	7.14	7.14	7.14	7.14	7.14
Pseudo R-squared	0.286	0.296	0.310	0.287	0.297	0.311
Panel C: All training, extensive definition, no threshold on the number of trained individuals						
	(13)	(14)	(15)	(16)	(17)	(18)
Wage growth differential	-1.215**	-1.214*	-1.367**	-1.923	-2.185	-2.232
	(0.589)	(0.625)	(0.600)	(1.529)	(1.636)	(1.546)
Residual from first-stage regression				0.745	1.023	0.914
				(1.517)	(1.603)	(1.520)
Basic controls	yes	yes	yes	yes	yes	yes
Controls for firm size	no	yes	yes	no	yes	yes
Extended controls	no	no	yes	no	no	yes
Number of observations	5393	5234	5057	5393	5234	5057
F-test on instrument significance	5.65	5.65	5.65	5.65	5.65	5.65
Pseudo R-squared	0.220	0.223	0.236	0.221	0.223	0.236

**Table 6 (cont.). Robustness checks. Dependent variable: Participation in general training in 1996.**

Panel D: All training, extensive definition, age 20-60

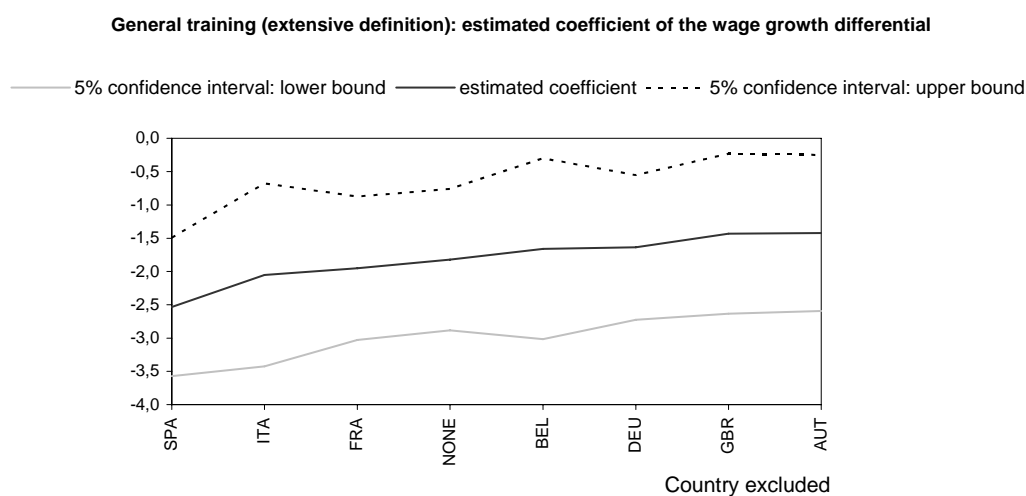
	(19)	(20)	(21)	(22)	(23)	(24)
Wage growth differential	-1.578** (0.702)	-1.562** (0.698)	-1.468** (0.689)	-2.702* (1.425)	-2.909* (1.488)	-2.616* (1.416)
Residual from first-stage regression				1.162 (1.363)	1.392 (1.450)	1.190 (1.352)
Basic controls	yes	yes	yes	Yes	yes	yes
Controls for firm size	no	yes	yes	No	yes	yes
Extended controls	no	no	yes	No	no	yes
Number of observations	7142	6929	6675	7142	6929	6675
F-test on instrument significance	5.91	5.91	5.91	5.91	5.91	5.91
Pseudo R-squared	0.182	0.188	0.204	0.183	0.188	0.204

Panel E: All training, extensive definition, age 20-50

	(25)	(26)	(27)	(28)	(29)	(30)
Wage growth differential	-1.273** (0.590)	-1.371** (0.597)	-1.355** (0.598)	-4.867** (2.036)	-5.252** (2.158)	-4.703** (2.048)
Residual from first-stage regression				3.565* (1.948)	3.843* (2.067)	3.315* (1.912)
Basic controls	yes	yes	yes	Yes	yes	yes
Controls for firm size	no	yes	yes	No	yes	yes
Extended controls	no	no	yes	No	no	yes
Number of observations	6046	5861	5638	6046	5861	5638
F-test on instrument significance	3.69	3.69	3.69	3.69	3.69	3.69
Pseudo R-squared	0.179	0.185	0.203	0.180	0.186	0.204

Note: Cluster-adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Individuals reporting firm-specific training are excluded from the sample. Each column of each panel refers to a different specification. The extensive definition of general training is used except in Panel B. In Panels A and B individuals reporting non-sponsored training are excluded from the sample (France is also excluded from the sample in this case since no French individual reported general employer-sponsored training). Each column of each panel refers to a different specification. In Panel C, the threshold of at least five trained individuals per cluster is removed. In Panels D and E, the age range of the sample is modified as indicated. Basic controls are country, education, occupation and sector dummies plus age and family variables. Specifications with extended controls add tenure, permanent job status and previous unemployment dummies to firm size dummies and basic controls. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. \*\*\*, \*\*, \*: significant at the 1%, 5% and 10% level of confidence, respectively.

Chart 1. Sensitivity to variation of country coverage



The figure shows central estimates and confidence intervals obtained by re-estimating the model (with extended controls) after excluding one country at a time from the sample. NONE identifies the specification with all 7 countries for the purpose of comparison.

## 5. Conclusions

44. According to Becker [1964], when labour markets are perfectly competitive, only the worker will invest in general training, since she is the only agent who can reap the benefits from the investment. To the extent that labour market imperfections are of limited empirical relevance as regards training, the Becker model represents a simple and powerful model to think about training in practical terms. We can test this statement by looking at the predictions of the model concerning the relationship between the training wage premium and training. In the Becker model, we expect that the greater the wage premium, the greater is the training investment. By contrast, as a large body of literature has pointed out, when labour markets are imperfectly competitive, firms may be willing to finance general training if the wage structure is compressed, that is if the increase of productivity after training is faster than the increase in pay. In this case, a negative relationship between the training wage premium and training might emerge, to the extent that the training wage premium is not positively correlated with wage compression.

45. In this paper, we contribute to the literature that tries to shed light on the empirical relevance of alternative theories of training, by exploiting the cross-country variation in training incidence and training wage premia in some European Union countries. We find that the probability of receiving general training, proxied in the paper with off-site training, is higher in clusters — defined by country, sector, occupation and educational attainment — with a lower differential between the median wage growth of trained and untrained employees. Importantly, the negative and statistically significant correlation between the training wage premium and training incidence does not seem to reflect the potential endogeneity of the former variable and holds also when general training is employer-provided.

46. While statistically significant, the estimated impact of changes in the training wage premium on the probability of general training is rather small: conditional on an extended set of controls, a 1% increase in the wage growth differential is expected to reduce the probability of general training by about 0.03%. However, to the extent that productivity gains from training and training wage premia are not negatively correlated, this number is likely to be only a lower bound estimate of the effect of wage compression on

general training. Should matched employer-employee data become available on a cross-country comparable basis, the magnitude of the effect of wage compression on training incidence could be precisely estimated using our empirical framework, by controlling for the productivity gains from training at the individual and firm level.

47. Overall, our findings suggest that competitive theories of the labour market, which imply a positive relationship between the training wage premium and the incidence of general training, provide insufficient guidance to interpret empirical training patterns. In particular, one should be cautious before assuming that labour market reforms inducing a greater alignment of wages and productivity would automatically lead to more investment in human capital. Conversely, training patterns are not inconsistent with the view that economic environments with higher wage compression can help firms organize and pay for general training, as predicted by the recent theories of training in imperfect labour markets.

## APPENDIX

### *A.1 Definition of co-variates and descriptive statistics*

48. *Wage growth differentials* are defined in terms of gross hourly wages, computed from normal gross monthly earnings in the main job at the date of the interview, by dividing them by 52/12 and by the number of usual weekly hours of work. Overtime pay and hours are included, but individuals who are either still in training in the final year or work less than 30 hours per week (despite declaring to work full-time) or more than 70 hours per week are excluded. Similarly, we exclude workers who change cluster between the two interviews (or with missing cluster affiliation in the first interview). Wage growth differentials (as well as log age differentials) are computed using ECHP personal weights. Although in principle there are 84 clusters, non-missing wage growth information for both individuals receiving training and not receiving training is available in 1996 only in 79 clusters (with an average training incidence of 15.1%).<sup>33</sup> Furthermore, the thresholds of 30 observations and five training events imply that the effective sample is reduced to 47 clusters, although, not surprisingly, retained clusters are larger and account for about 90% of observations with non-missing wage growth information (with an average training incidence of 14.9%).<sup>34</sup>

49. As described in the text, all specifications control for age, educational attainment, occupation, sector and country as well as for a set of family variables. Family variables (see Table A2) include household type (grouped into eight categories), a dummy for presence of children aged less than 12 years, a dummy that equals 1 if the respondent thinks that looking after other persons prevented him from working as much as he wanted, and a series of variables proxying the financial situation of the family (including indicators of home ownership, the burden of accommodation costs including mortgage, the number of rooms divided by household size and whether the household received an heritage of at least EUR 2 000 in the last 12 months). In the regressions, age is measured in years, while all other co-variates (with the exception of wage growth differentials and rooms per household member) are categorical, and coded as sets of dummy variables (omitting one dummy per set for identification). We consider three educational attainment levels (less than upper secondary, upper secondary, more than upper secondary, corresponding to standard ISCED categories 0-2, 3, and 5-7), eight occupational groups (corresponding to standard ISCO-88 codes 1 to 9 with the exclusion of skilled agricultural workers, or ISCO-88 code 6), 13 sectors (listed in Table A1), 4 firm-size classes, a dummy for permanent contract, a dummy for at least one spell of unemployment since 1989, and five tenure classes. Tenure is obtained as the difference between the survey year and the calendar year of start of the current job. We grouped the data into five classes rather than using a continuous variable for two reasons: *i*) the information is censored at 15 years; and *ii*) computed this way, tenure measures are highly imprecise. For example, an individual who is surveyed in December but was hired in January of the survey year would be accounted as having tenure shorter than

---

33. In the case of Austria, wage growth is computed for 1997, due to data quality problems (see Section A.2 in this appendix).

34. Training incidence in the sample used to compute wage growth differentials is smaller than in the sample used for the final regressions because individuals still in training at the time of the interview are excluded in the former but obviously not in the latter.

another individual who started in December of the year before the survey year but was surveyed in January.<sup>35</sup> The means of the co-variates used in the regressions are reported in Table A2.<sup>36</sup>

**Table A1. List of included sectors**

NACE Rev.1 Codes	Sector definition
C+E	Mining and quarrying + Electricity, gas and water supply
DA	Manufacture of food products, beverages and tobacco
DB+DC	Manufacture of textiles, clothing and leather products
DD+DE	Manufacture of wood and paper products; publishing and printing
DF-DI	Manufacture of coke, refined petroleum/chemicals/rubber & plastic products
DJ+DK	Manufacture of metal products, machinery and equipment n.e.c.
DL-DN	Other manufacturing
F	Construction
G	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal/household goods
H	Hotels and restaurants
I	Transport, storage and communication
J	Financial intermediation
K	Real estate, renting and business activities

Note: excluded sectors are: agriculture (A+B) as well as public administration and defence, compulsory social security, education, health and social work, other community, social and personal service activities, private households with employed persons and extra-territorial organizations and bodies (L to Q).

### *A.2 Data quality and sample selection*

50. Our ECHP release contains longitudinal data for five years (1994-1998), although not all countries are available in each wave. The country sample is chosen on the basis of data availability: Sweden has no wage growth data; Greece, Ireland, Luxembourg and Portugal, have either a very small sample or very small training incidence, so that almost no cluster would be retained for these countries after applying the thresholds of 30 observations and 5 training events; and Denmark, the Netherlands and Finland have no information on training type (or that information is almost always missing). Thus, these eight countries are excluded from the sample, leaving us with seven countries (Austria, Belgium, France, Germany, Italy, Spain, and the United Kingdom), although there are no data for Austria in 1994 and complete data for France, Germany and the UK are available only until 1996 included (for the UK the disaggregation by type of training is however available only in 1996).

35. Information on the month of the interview is also available in the ECHP. However, this information is always missing in Germany (for confidentiality reasons) and often missing in a few other countries. For this reason, we opted not to use it.

36. The standard deviations of age and rooms per household member, the only continuous variables in the sample beside the wage growth differential, are 8.1 years and 0.74, respectively.

**Table A2. Means of co-variates (baseline sample: age 30 to 60 years, year: 1996).**

<i>Variable</i>	<i>Mean</i>	<i>Variable</i>	<i>Mean</i>
Age (years)	42.2	% NACE G	14.2
% tenure 1 year or less	12.6	% NACE H	2.4
% tenure 2 to 5 years	16.4	% NACE I	6.6
% tenure 6 to 9 years	15.1	% NACE J	6.5
% tenure 10 to 14 years	11.5	% NACE K	6.9
% tenure 15 years or more	44.4	% Austria	10.9
% firm size less than 50 employees	45.1	% Belgium	8.0
% firm size 50-99 employees	10.5	% France	18.8
% firm size 100-499 employees	18.5	% Germany	16.2
% firm size 500 employees or more	25.9	% Italy	17.6
% tertiary education	19.1	% Spain	18.4
% upper secondary education	43.3	% United Kingdom	10.1
% less than upper secondary education	37.6	<b>Family variables:</b>	
% Legislators, senior officials and managers	9.3	Number of rooms per household member	1.39
% Professionals	7.1	% received an heritage in the last 12 months	1.98
% Technicians and associate professionals	12.8	% housing costs are a heavy burden	19.55
% Clerks	10.6	% housing costs are somewhat a burden	46.45
% Service and shop and market sales workers	5.5	% housing costs are not a problem	33.99
% Craft and related trades workers	31.0	% house owners	71.43
% Plant and machine operators and assemblers	16.0	% tenants	24.59
% Elementary occupations	7.8	% with rent-free accommodation	3.98
% previous unemployment	29.3	% limited in work effort by looking after others	0.95
% permanent contract	89.7	% children aged less than 12 in the household	43.73
% NACE C+E	2.8	% one person household	5.92
% NACE DA	4.8	% single parent with one or more children	3.10
% NACE DB+DC	2.7	% couple without children	14.96
% NACE DD+DE	4.7	% couple with one child (aged less than 16)	13.06
% NACE DF-DI	7.9	% couple with two children (all children aged less than 16)	17.51
% NACE DJ+DK	14.2	% couple three or more children (all children aged less than 16)	5.43
% NACE DL-DN	11.1	% couple with children (at least one child aged 16 years or more)	33.46
% NACE F	15.2	% other households	6.56

51. Unit non-responses and attrition rates in the ECHP are comparable with those of other longitudinal household surveys [Peracchi, 2002]. Nevertheless, due to small entry rates, attrition results in a reduction of the sample size that is increasing with time. Focusing only on the countries that are present in two consecutive waves, the loss of observations due to attrition and insufficient replacement amounts to about 7% between the first (1994) and the second (1995) wave, to 4% between the second and the third (1996) wave, and to 6% between the third and the fourth (1997) wave, as well as between the fourth and the fifth (1998) wave.

52. We choose to undertake a cross-sectional analysis of 1996 data only (and not to exploit further the longitudinal structure of the ECHP by pooling together more than one wave) because this year has much larger country coverage, higher number of clusters, smaller non-response rates and bigger sample size. In fact, in 1996 we can cover seven countries in our data. Conversely, taking into account the lack of data on training type in certain countries and years, we can include only six countries (Austria, Belgium, France, Germany, Italy and Spain), in 1995, four countries (Austria, Belgium, Italy and Spain), in 1997,

and, taking into account item non-responses, only three countries (Austria, Belgium and Spain), in 1998. After applying our size thresholds, there are 47 clusters in 1996,<sup>37</sup> while there can be at most 40 clusters in 1995,<sup>38</sup> 22 clusters in 1997 and only 16 clusters in 1998. Moreover, in 1996 there are 7 196 male wage and salary employees, aged from 30 to 60 years and working full-time in the non-agricultural market-based sectors. Excluding observations with missing values either in the training variables or other covariates (with the exception of the wage growth differential), this sample drops to 6 230 individuals, corresponding to 87% of the original sample. If we also drop observations for which the wage growth differential is not defined (because of the thresholds of 30 observations and five trained individual per cluster), the sample drops further to 5 515 observations (77% of the original sample). Conversely, male wage and salary employees, aged from 30 to 60 years and working full-time in the non-agricultural market-based sectors are 7 618, 3 827 and 2 546 in 1995, 1997 and 1998, respectively. These figures drop to only 3 896, 3 607 and 2 407, respectively, when observations with missing values are excluded, and drop even further when observations for which the wage growth differential is not defined are excluded.

53. We also perform a sensitivity analysis on 1995 data (see below). With 1995 data, however, we are confronted with an additional problem: in many countries, wage data from the first available ECHP wave are not comparable with data from subsequent waves (see OECD [2004]). For this reason, we use 1996-1997 wage growth rates for Austria – the country in our sample with data starting from 1995 - throughout the analysis and we use 1995-1996 wage growth rates in the sensitivity analysis with 1995 training data, for all countries except Austria.

54. Eurostat's guidelines specify that National Data Collection Units (NDUs) are requested to use country-specific categories of education and training according to the classification used in the national Labour Force Surveys (LFS). In order to facilitate the coding of training questions by the NDUs, Eurostat provides a correspondence between LFS and ECHP categories. In the case of France and the United Kingdom, no national LFS category corresponds to ECHP category d (specific vocational training in a working environment, without complementary instruction elsewhere). However, some training is reported to fall under this category in the British and French data. We choose to keep these countries in our sample, but we also check whether our findings are driven by misclassification by replicating our regression analysis on a sub-sample of countries which exclude France and the UK. All results that are presented in the paper are robust to the simultaneous elimination of these two countries (detailed results are available from the authors on request).

### *A.3 Full estimation results for Table 4*

---

37. 43 without Austria, for which we use 1996-1997 wage growth differentials (see below).

38. 36 without Austria.

Table A3. *Probits for total training*. Dependent variable: training participation *T* in 1996.

	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.019*	-0.022*	-0.024*	-0.019*	-0.022*	-0.024*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Educational attainment (ref: More than upper secondary)						
Upper secondary education	-0.233**	-0.247**	-0.280**	-0.253**	-0.271**	-0.299***
	(0.104)	(0.113)	(0.113)	(0.104)	(0.112)	(0.111)
Less than upper secondary education	-0.461***	-0.422***	-0.471***	-0.475***	-0.438***	-0.485***
	(0.117)	(0.136)	(0.134)	(0.113)	(0.131)	(0.129)
Previous unemployment spell			-0.140**			-0.139**
			(0.061)			(0.061)
Tenure (ref: 0-1 years)						
Tenure: 2-5 years			-0.235**			-0.234**
			(0.117)			(0.117)
Tenure: 6-9 years			-0.236**			-0.238**
			(0.103)			(0.103)
Tenure: 10-14 years			-0.133			-0.134
			(0.102)			(0.102)
Tenure: 15 years or more			-0.178			-0.177
			(0.113)			(0.113)
Permanent job			0.259***			0.259***
			(0.096)			(0.095)
Firm size (ref: 1-49 employees)						
Firm size: 50-99		0.206**	0.194**		0.208**	0.197**
		(0.081)	(0.082)		(0.081)	(0.081)
Firm size: 100-499		0.396***	0.342***		0.402***	0.349***
		(0.074)	(0.070)		(0.073)	(0.069)
Firm size: 500 or more		0.492***	0.466***		0.497***	0.471***
		(0.073)	(0.082)		(0.072)	(0.081)
Wage growth differential	-0.193	-0.168	-0.345	-1.890	-2.318*	-2.386**
	(0.411)	(0.464)	(0.417)	(1.295)	(1.319)	(1.219)
Residual from first-stage regression				1.777	2.254	2.148
				(1.393)	(1.450)	(1.344)
Joint significance of family variables (P-value)	0.001	0.005	0.006	0.001	0.005	0.006
Number of observations	5898	5728	5515	5898	5728	5515
Pseudo R-squared	0.154	0.171	0.179	0.154	0.171	0.179

Note: Cluster-adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Additional controls not in the table are family variables and country, occupation and sector dummies. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. \*\*\*, \*\*, \*: significant at the 1%, 5% and 10% level of confidence, respectively.

#### A.4 Further sensitivity analysis

55. A feature of the analysis in the main text is that we have ignored firm – specific training. In Table A4, we take more explicitly into account the fact that general and specific training are mutually exclusive categories by design by estimating a multinomial logit model with three outcomes, using no training as the baseline category. In the case of total training, the dependent variable (*TYPE*) takes the value 0 in the case of no training, 1 in the case of general training, and 2 in the case of firm-specific training. In the case of employer-sponsored training, we define another dependent variable (*TYPEM*), which takes value 0 in the case of no training, 1 in the case of employer-sponsored general training, 2 in the case of employer-sponsored firm-specific training, and 3 in the case of non-sponsored training. General and firm-specific non-sponsored training are grouped together due to the small number of non-sponsored training events in our sample. As before, we use both the “extensive” and the “restrictive” definition of general training. We have also estimated extended specifications that include the residual from the first-step regression, which turned out to be always statistically insignificant (with confidence intervals obtained by bootstrapping, using 1000 i.i.d. resamplings). Results are available upon request.

56. We find that the estimated coefficient associated to our measure of the training wage premium  $\Delta W_c$  is always negative and statistically significant, in the case of general training, and close to zero and not statistically significant, in the case of firm–specific training. Note, however, that the latter finding cannot be taken as evidence that wage compression does not provide firms with an incentive for firm-specific training. The reason is that if the true premium to firm-specific training is much smaller than the true premium to general training, our measure of the wage premium will tend to be more correlated with the latter than with the former. In other words, our simple measure is too noisy a proxy of the wage premium associated to firm-specific training to allow us to infer something about the relationship between the wage structure and firm-specific training.

57. We explore also the effect of the wage growth differential on each separate category of training. In order to do so, we define two new dependent variables: *DTYPE*, which takes value 0 in the case of no training, 1 in the case of third level qualification or other vocational training at a vocational school or college (we group together these two categories due to the small number of training spells leading to third level qualification in the sample), 2 in the case of dual systems with both workplace and school training and 3 in the case of workplace training only; and *DTYPEM* which is equal to *DTYPE* for employer-sponsored training and to 4 when training is not employer-provided. Since these categories are not naturally ordered, but are mutually exclusive by design, we estimate two multinomial logit models with 4 and 5 categories, respectively, using no training as the baseline category.

58. Table A5 shows the estimated coefficients associated to the wage growth differential in the specifications with extended controls (similar results are obtained if we use basic controls). Point estimates are higher for training which is at least partially off-site – categories 1 and 2 – than for workplace training – category 3 – although standard errors of the second category are relatively large, due probably to the small number of training spells falling into it.

**Table A4. General and specific training. Trinomial and quadrinomial logits. Dependent variable: Total and employer provided training in 1996.**

Panel A: Extensive definition of general training

	(1)	(2)	(3)	(4)
	<i>Total training</i>		<i>Employer-sponsored training</i>	
	<i>Basic controls</i>	<i>Extended controls</i>	<i>Basic controls</i>	<i>Extended controls</i>
General training	-3.945*** (1.218)	-4.168*** (1.150)	-4.307*** (1.458)	-4.946*** (1.412)
Firm-specific training	-.085 (.962)	-.425 (1.028)	.251 (1.010)	-.247 (1.067)

Panel B: Restrictive definition of general training.

	(5)	(6)	(7)	(8)
	<i>Total training</i>		<i>Employer-sponsored training</i>	
	<i>Basic controls</i>	<i>Extended controls</i>	<i>Basic controls</i>	<i>Extended controls</i>
General training	-3.853** (1.717)	-4.121** (1.723)	-4.146** (1.962)	-4.747** (1.940)
Firm-specific training	-.412 (.864)	-.772 (.951)	-.135 (.977)	-.660 (1.053)

Note: Cluster-adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Each column of each panel refers to a different specification. In Columns (1)-(2) and (5)-(6) the dependent variable is the type of training (TYPE), while in Columns (3)-(4) and (7)-(8) the dependent variable is the type of employer-sponsored training (TYPEM). Only estimated coefficients of the wage growth differential are reported. Basic controls are country, education, occupation and sector dummies plus age and family variables. Extended controls add tenure, firm size, permanent job status and previous unemployment dummies. \*\*\*, \*\*, \*: significant at the 1%, 5% and 10% level of confidence, respectively.

**Table A5. Disaggregating training type. Dependent variable: Total and employer provided training. Multinomial logits with 4 and 5 characteristics, respectively.**

	(1)	(2)
	<i>Total training</i>	<i>Employer provided training</i>
1. vocational training at school or college	-4.202** (1.726)	-4.970** (1.933)
2. workplace training and complementary instruction elsewhere	-2.373 (2.348)	-3.405 (3.008)
3. workplace training without complementary instruction elsewhere	-.421 (1.029)	-.260 (1.067)

Note: Cluster-adjusted robust standard errors within parentheses. In Column (1) the dependent variable is the disaggregated type of training (DTYPE), while in Column (2) the dependent variable is the disaggregated type of employer-sponsored training (DTYPEM). Only estimated coefficients of the wage growth differential are reported. Observations are weighted by ECHP personal weights. Other controls are country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies plus age and family variables. \*\*\*, \*\*, \*: significant at the 1%, 5%, 10% level of confidence, respectively.

59. In Table A6, we consider a number of other robustness checks using the probit specification with extended controls and the extensive definition of general training. First, we try to control better for composition effects by adding to our regressors country-by-sector dummies and the employment rate in 1996 by country and educational attainment (which we matched to our sample using data from the European Labour Force Survey).<sup>39</sup> Second, we modify cluster definitions by dividing the group of

39. One reason to include country-by-industry dummies is that the composition of each sector varies by country, which could bias our estimates if better workers flow into sub-sectors where training opportunities and wage prospects are better. We also include the employment rate, which varies by country and educational attainment, because greater statutory or contractual minimum wages may reduce employment among the

medium/low-skilled workers into low-skilled and medium-skilled workers. This is equivalent to excluding the former category, since training incidence is small in that group (see Table 1). We also consider three alternative splits of the group of medium/low-skilled occupations: *i*) between clerks and all the other groups; *ii*) between white collars and sales workers, on the one hand, and other blue collars, on the other; and *iii*) between white collars, sales workers and craft and related trade workers, on the one hand, and plant and machine operators and assemblers and low-skilled occupation, on the other. Third, we check whether our results are affected by the fact that sectors where non-profit organizations are not important are included in the computation of the wage growth differential but not in the final regressions, by estimating the baseline specification on all sectors or excluding these sectors from the computation of  $\Delta W_c$ . Fourth, we check the robustness of our results to measurement error in five additional ways: i) we compute  $\Delta W_c$  by defining as trained only those workers who started and ended the training course between two subsequent interviews;<sup>40</sup> ii) we compute  $\Delta W_c$  including also workers who change cluster between the two interviews; iii) we compute  $\Delta W_c$  only for workers who did not change employer between the two interviews; iv) we use wage level differentials instead of wage growth differentials to measure training wage premia; and v) we compute  $\Delta W_c$  as the difference between the median wage growth rates of workers who took *general* training and of those who took no training. Finally, we lift both our size thresholds of 30 observations and 5 trained individuals per cluster.

60. In the last two rows of the table, we also verify whether our results can be attributed to the selected year (1996) or to the choice of the year for the computation of the wage growth differential (1995-1996, except for Austria). First, we re-estimate our models on 1995 data (except for wage growth differentials).<sup>41</sup> Second, we take the average of the wage growth differentials in 1995-1996 and 1996-1997 for the countries where both differentials can be computed and we re-estimate our models by using this average as our measure of  $\Delta W_c$ .

61. Overall, from these sensitivity exercises it can be concluded that our key findings are robust to changes in the definition of the key variables and in the model specification.

---

untrained (particularly among the least able workers) without affecting the number of workers who are trained, thereby driving up training incidence among the employed and driving down the wage growth differential (we also experimented with education-by-country dummies with similar results).

40. The ECHP contains start and end dates of the last training spell that, in principle, could be matched with the date of the previous interview to exactly identify those individuals who took some training between two interviews. In practice, however, this information is often missing. Furthermore, interview dates are not available for Germany. For this reason, to preserve sample size and country coverage, we prefer to compute wage growth differentials using exact training dates only for this sensitivity analysis.
41. Wage data in 1994 are not comparable with data in 1995. For this reason, wage growth differentials for 1994-1995 cannot be computed (see above).

**Table A6. Other robustness tests. Probit models. Dependent variable: Participation in general training (extensive definition).**

A. Country X sector dummies	-1.599** (0.695)	I. Exact training dates to compute $\Delta W_c$	-2.101*** (0.799)
B. Country X sector dummies + empl. rate	-1.399* (0.742)	J. Cluster-movers also to compute $\Delta W_c$	-1.721*** (0.549)
C. No low skilled	-1.854*** (0.609)	K. Job-stayers only to compute $\Delta W_c$	-1.500*** (0.581)
D. Three occupational groups (1)	-1.124* (0.646)	L. Wage level data to compute $\Delta W_c$	-0.576* (0.302)
E. Three occupational groups (2)	-0.829 (0.622)	M. $\Delta W_c$ based on general training only	-2.175*** (0.437)
F. Three occupational groups (3)	-1.749** (0.694)	N. No size thresholds	-0.730** (0.335)
G. All sectors	-1.212** (0.532)	O. Wave 2 (1995)	-2.047* (1.102)
H. No non-market sectors to compute $\Delta W_c$	-1.426*** (0.442)	P. Wave 3 (1996), average $\Delta W_c$	-1.109** (0.567)

Note: Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Individuals reporting firm-specific training are excluded from the sample. Each cell of the table refers to a different regression. Only estimated coefficients of the wage growth (or wage level) differential are reported. General training is defined according to the extensive definition. Other controls are country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies plus age and family variables. In Row A country by sector dummies are added to the baseline specification; in Row B both country by sector dummies and the employment rate (by country and education) are added to the baseline specification; in row C elementary occupations are excluded; in row D the group of medium/low-skilled occupations is divided into two groups (clerks, and other medium/low skilled workers) and clusters are defined accordingly; in row E service workers and shop and market sales workers are switched from the group of other medium/low skilled workers to the group of clerks; in row F craft and related trades workers are switched from the group of other medium/low skilled workers to the group of clerks, service workers and shop and market sales workers; in row G sectors with NACE Rev.1 code greater than K are included in the sample; in row H sectors with NACE Rev.1 code greater than K are not used in the computation of the wage growth differential; in row I wage growth differentials are computed by defining as trained only workers who started training after the date of the 1995 interview (1996 interview for Austria); in row J wage growth differentials are computed by including workers who changed cluster between the two interviews (or with missing cluster affiliation in the first interview) in which wages are measured; in row K wage growth-differentials are computed by excluding workers who changed employer between the two interviews in which wages are measured; in row L wage level differentials in 1996 are used as proxy of the training premium; in row M general training rather than total training is used to compute the wage growth differential; in row N the size threshold (30 observations per cluster, of which at least 5 receiving training) is removed; in Row O the baseline specification is estimated using data from 1995 (1995-1996 for the wage growth differential); in row P the baseline specification is estimated using data from 1996 (but the average of 1995-1996 and 1996-1997 for the wage growth differential). \*\*\*, \*\*, \*: significant at the 1%, 5%, 10% level of confidence, respectively.

**A.5 Proof that if younger workers are trained first, and the log age distribution has the same degree of concentration in the neighbourhood of the two medians, the log age differential does not depend on the aggregate training rate (statement made in footnote 13).**

62. Let us assume that all workers with log age  $x < a(\tau)$  are trained, where  $\tau$  is the training rate and  $da/d\tau > 0$ . The median of the log age distribution of the trained  $M_T$  is implicitly defined by:

$$\frac{\int_{-\infty}^{M_T} f(x)dx}{\int_{-\infty}^a f(x)dx} = .5, \text{ which implies } \int_{-\infty}^{M_T} f(x)dx = .5 \int_{-\infty}^a f(x)dx \text{ and, differentiating with respect to } a \text{ (which is}$$

the only parameter that depend on  $\tau$ ):

$$f(M_T) \frac{dM_T}{da} = .5 f(a). \quad [\text{A1}]$$

Repeating the same reasoning for the untrained, we have:

$$\frac{\int_a^{M_U} f(x)dx}{\int_a^{\infty} f(x)dx} = .5, \text{ which yields } 1 - \int_{-\infty}^{M_U} f(x)dx = .5 - .5 \int_{-\infty}^a f(x)dx \text{ and differentiating:}$$

$$f(M_U) \frac{dM_U}{da} = .5 f(a). \quad [\text{A2}]$$

Combining [A1] with [A2], we obtain  $f(M_U) \frac{dM_U}{da} = f(M_T) \frac{dM_T}{da}$ . Now, if  $f(M_U) = f(M_T)$ , an hypothesis that we do not reject when using the log age (see Table 3 in the text), then  $\frac{dM_U}{da} = \frac{dM_T}{da}$  and therefore  $\frac{d(M_U - M_T)}{da} = 0$ . In other words, if we increase the proportion of trained workers, the log age differential  $M_U - M_T$  will not be affected, since it is not affected by either the log age of switching individuals  $a$ , or the fraction of switching individuals in the population  $f(a)$ , or the exact values of the medians  $M_T$  and  $M_U$ .

## REFERENCES

- Abowd, J.M. and F. Kramarz (2003), “The cost of hiring and separations”, *Labour Economics*, 10, 499-530.
- Acemoglu, D. and J. Pischke (1999a), “The structure of wages and investment in general training”, *Journal of Political Economy*, 107, 539-572.
- Acemoglu, D. and J. Pischke (1999b), “Beyond Becker: Training in Imperfect Labour Markets”, *The Economic Journal*, 109, F112-F142.
- Acemoglu, D. and J. Pischke (2003), “Minimum Wage and On-the-Job Training”, *Research in Labor Economics*, 23, 159-202.
- Arulampalam, W., A. Booth and M. Bryan (2004), “Training and the New Minimum Wage”, *The Economic Journal*, 114, C87-C94.
- Aubert P. and B. Crépon (2004), « La productivité des salariés âgés: une tentative d’estimation », *Economie et Statistique*, 368, 95-119.
- Barron, J.M., M.C. Berger and D.A. Black (1999), “Do Workers Pay for On-the-Job Training?”, *Journal of Human Resources*, 34, 235-252.
- Bartelsman, E., S. Scarpetta and F. Schivardi (2005), “Comparative Analysis of Firm Demographics and Survival: Micro-level Evidence for the OECD Countries”, *Industrial and Corporate Change*, forthcoming.
- Bassanini, A and G. Brunello (2003), “Is training more frequent when wage compression is higher? Evidence from the European Community Household Panel”, IZA Discussion Paper 839.
- Becker, G. (1964), *Human Capital*, NBER, New York.
- Bishop, J.H. (1997), “What We Know about Employer-Provided Training: A Review of the Literature”, *Research in Labor Economics*, 16, 19-87.
- Blundell, R., L. Dearden and C. Meghir (1999), “Work-related training and earnings”, IFS working paper, Institute for Fiscal Studies, London.
- Booth, A and M. Bryan (2005), “Testing some predictions of human capital theory: New training evidence from Britain”, *Review of Economics and Statistics*.
- Booth, A. and G. Zoega (2004), “Is wage compression a necessary condition for firm-financed general training?”, *Oxford Economic Papers*, 56, 88-97.
- Frazis, H. and M. Loewenstein (1999), “Re-examining the Returns to Training: Functional Form, Magnitude, and Interpretation”, BLS Working Paper 325.
- Gerfin, M. (2004), “Firm-Sponsored General Training in Frictional Labour Markets: An Empirical Analysis for Switzerland”, IZA Discussion Paper 1077.

- Grossberg, A.J. and P. Sicilian (1999), "Minimum Wages, On-the-Job Training, and Wage Growth", *Southern Economic Journal*, 65, 539-556.
- Hashimoto, M. (1981), "Firm-Specific Human Capital as a Shared Investment", *American Economic Review*, 71, 475-482.
- Katz, E. and A. Ziderman (1990), "Investment in general training: the role of information and labor mobility", *The Economic Journal*, 100, 1147-1158.
- Lazear, E.P. (2003), "Firm-Specific Human Capital: A Skill-Weights Approach", NBER Working Paper, No. 9679, May.
- Lengermann, P. (1999), "How Long Do the Benefits of Training Last? Evidence from Long Term Effects across Current and Previous Employers", *Research in Labor Economics*, 18, 439-461.
- Loewenstein, M and J. Spletzer (1999), "General and Specific Training: Evidence and Implications", *Journal of Human Resources*, 34, 710-733.
- Loewenstein, M. and J. Spletzer (1998), "Dividing the costs and the returns to general training", *Journal of Labor Economics*, 16, 142-171.
- Moen, E. and A. Rosen (2002), "Does Poaching Distort Training?", CEPR Discussion Paper, No. 3468, July.
- Moulton, B. (1986), "Random Group Effects and the Precision of Regression Estimates", *Journal of Econometrics*, 32, 385-397.
- Nestler, K. and E. Kailis (2002), "Disparities in Access to Continuing Vocational Training in Enterprises in Europe", Eurostat Statistics in Focus No. 22/2002.
- Neumark, D. and W. Wascher (2001), "Minimum Wages and Training Revisited", *Journal of Labor Economics*, 19, 563-595.
- OECD (2003), Employment Outlook.
- OECD (2004), Employment Outlook.
- Peracchi, F. (2002), "The European Community Household Panel: A Review", *Empirical Economics*, 27, 63-90.
- Pischke, J.-S. (2001), "Continuous Training in Germany", *Journal of Population Economics*, 14, 523-548.
- Schøne, P. (2004), "Why Is the Return to Training So High?", *Labour*, 18, 363-378.
- Sicilian, P. (2001), "On-the-Job Training and Starting Wages", *Journal of Labor Research*, 22, 809-816.
- Smith, R. and R. Blundell (1986), "An exogeneity test for a simultaneous equation tobit model with an application to labor supply", *Econometrica*, 54, 679-685.
- Stevens, M. (1994), "A theoretical model of on-the-job training with imperfect competition", *Oxford Economic Papers*, 46, 537-562.

**OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS**

Most recent releases are:

- No. 40 *NEUTRAL OR FAIR? ACTUARIAL CONCEPTS AND PENSION-SYSTEM DESIGN (2006)*  
Monika Queisser and Edward Whitehouse
- No. 39 *STARTING WELL OR LOSING THEIR WAY? THE POSITION OF YOUTH IN THE LABOUR MARKET IN OECD COUNTRIES (2006)*  
Glenda Quintini and Sébastien Martin
- No. 38 *SOCIAL ASSISTANCE POLICY DEVELOPMENT AND THE PROVISION OF A DECENT LEVEL OF INCOME IN SELECTED OECD COUNTRIES (2006)*  
Willem Adema
- No. 37 *MEASURES OF MATERIAL DEPRIVATION IN OECD COUNTRIES (2006)*  
Romina Boarini and Marco Mira d'Ercole
- No. 36 *FROM INACTIVITY TO WORK: THE ROLE OF ACTIVE LABOUR MARKET POLICIES (2006)*  
Stéphane Carcillo and David Grubb
- No. 35 *EMPLOYMENT PATTERNS IN OECD COUNTRIES: REASSESSING THE ROLE OF POLICIES AND INSTITUTIONS (2006)*  
Andrea Bassanini and Romain Duval
- No. 34 *AN AGE PERSPECTIVE ON ECONOMIC WELL-BEING AND SOCIAL PROTECTION IN NINE OECD COUNTRIES (2006)*  
Thai-Thanh Dang, Herwig Immervoll, Daniela Mantovani, Kristian Orsini, Holly Sutherland
- No. 33 *ALTERNATIVE MEASURES OF WELL-BEING (2006)*  
Romina Boarini, Asa Johansson and Marco Mira d'Ercole
- No. 32 *SOCIAL DISADVANTAGE AND EDUCATION EXPERIENCES (2006)*  
Stephen Machin
- No. 31 *CAN PARENTS AFFORD TO WORK? CHILDCARE COSTS, TAX-BENEFIT POLICIES AND WORK INCENTIVES (2005)*  
Herwig Immervoll and David Barber
- No. 30 *CHALLENGES FACING LABOUR OFFICES AND SOCIAL INSURANCE (2005)*  
Anders Reuterswärd
- No. 29 *NET SOCIAL EXPENDITURE, 2005 EDITION*  
Willem Adema & Maxime Ladaïque
- No. 28 *WELFARE REFORM IN EUROPEAN COUNTRIES: A MICROSIMULATION ANALYSIS (2005)*  
Herwig Immervoll, Henrik Jacobsen Kleven, Claus Thustrup Kreiner and Emmanuel Saez
- No. 27 *TRENDS AND DETERMINANTS OF FERTILITY RATES: THE ROLE OF POLICIES (2005)*  
Cristina d'Addio & Marco Mira d'Ercole
- No. 26 *THE "ENABLING STATE?" (2005)*  
Neil Guilbert
- No. 25 *COUNTING IMMIGRANTS AND EXPATRIATES IN OECD COUNTRIES: A NEW PERSPECTIVE (2005)*  
Jean-Christophe Dumont and Georges Lemaître
- No. 24 *TAXATION, ETHNIC TIES AND THE LOCATION CHOICE OF HIGHLY SKILLED IMMIGRANTS (2005)*  
Thomas Liebig and Alfonso Sousa-Poza

Recent available working papers can be found on the OECD website: <http://www.oecd.org/els/workingpapers>.

Other series of working papers available from the OECD include: OECD HEALTH WORKING PAPERS

**RECENT RELATED OECD PUBLICATIONS:**

*SICKNESS, DISABILITY AND WORK: BREAKING THE BARRIERS (VOL. 1) – NORWAY, POLAND AND SWITZERLAND (2006)*

*INTERNATIONAL MIGRATION OUTLOOK: SOPEMI - 2006 Edition*

*AGEING AND EMPLOYMENT POLICIES: LIVE LONGER, WORK LONGER (2006)*

*BABIES AND BOSSES: Reconciling Work and Family Life, Volume 4 Canada, Finland, Sweden and the United Kingdom (2005)*

*PENSIONS AT A GLANCE: Public policies across OECD countries (2005)*

*EXTENDING OPPORTUNITIES – How active social policy can benefit us all (2005)*

*SOCIETY AT A GLANCE: OECD Social Indicators (2005)*

*OECD EMPLOYMENT OUTLOOK (2006)*

*AGEING AND EMPLOYMENT POLICIES: AUSTRALIA (2005)*

*AGEING AND EMPLOYMENT POLICIES: AUSTRIA (2005)*

*AGEING AND EMPLOYMENT POLICIES: CANADA (2005)*

*AGEING AND EMPLOYMENT POLICIES: DENMARK (2005)*

*AGEING AND EMPLOYMENT POLICIES: FRANCE (2005)*

*AGEING AND EMPLOYMENT POLICIES: GERMANY (2005)*

*AGEING AND EMPLOYMENT POLICIES: IRELAND (2006)*

*AGEING AND EMPLOYMENT POLICIES: NETHERLANDS (2005)*

*AGEING AND EMPLOYMENT POLICIES: UNITED STATES (2005)*

*PROMOTING ADULT LEARNING (2005)*

*MIGRATION, REMITTANCES AND DEVELOPMENT (2005)*

*MIGRATION FOR EMPLOYMENT: BILATERAL AGREEMENTS AT A CROSSROADS (2004)*

*INCOME DISPARITIES IN CHINA: AN OECD PERSPECTIVE (2004)*

*BENEFITS AND WAGES: OECD Indicators (2004)*

*REFORMING PUBLIC PENSIONS: SHARING THE EXPERIENCES OF TRANSITION AND OECD COUNTRIES (2004)*

*BABIES AND BOSSES: Reconciling Work and Family Life, Volume 3, New Zealand, Portugal, Switzerland (2004)*

*ASSET BUILDING AND THE ESCAPE FROM POVERTY: A NEW WELFARE POLICY DEBATE (2003)*

*MANAGING DECENTRALISATION: A NEW ROLE FOR LABOUR MARKET POLICY (2003)*

*COMBATING CHILD LABOUR: A REVIEW OF POLICIES (2003)*

For a full list, consult the OECD On-Line Bookstore at [www.oecd.org](http://www.oecd.org)