On the Link Between On-the-Job Training and Earnings' Dispersion

Said Hanchane

Institut d'Economie Publique (IDEP), Marseille, and Laboratoire d'Economie et de Sociologie du Travail (LEST), Aix-en-Provence, France

and

Jacques Silber

Department of Economics, Bar-Ilan University, Israel

1. Introduction

Importance of a skilled workforce:

- **at the individual level: higher levels of human capital increase earnings and the probability of finding a job**
- **at a more aggregated level: a more competent labor force improves the chances of success of firms as well as of nations.**

Human capital investments however do not include only education but also continuing vocational training.

Several studies (e.g. OECD,1999) stressed the fact that

- training reinforces skill differences resulting from unequal participation in schooling
- workers receive more training in countries with higher average levels of education
- the wage premium associated with training differs between educational and gender groups

Béret and Dupray (2000) emphasized additional aspects of this unequal access to training such as the fact that training seems to be positively correlated with

- the professional status of the individual in the firm
- the nature of his work contract
- the size of the firm
- seniority in the firm

Such conclusions may however imply that the main goal of training is not to increase productivity but to “keep the workers in the firm”.

In fact (see, OECD, 1999) earnings growth after training and the event of training may not be independent variables so that unobserved individual characteristics may determine both

- the probability that someone is trained
- the fact that they earn higher-than-average wages after the training.

The OECD study thus concluded that

- half of the earnings gap between those who received training and those who did not is due to the fact that firms providing training pay higher salaries in any case
- the second half of the gap is related to factors that have a simultaneous impact on the probability of access to training and on earnings

The impact of on-the-job training on earnings should therefore be computed, net of the effect of this unobserved heterogeneity.

Assume we find

- first that there is such a selectivity bias
- second that there remains a net (of the role played by the unobserved heterogeneity) effect of on-the-job training on earnings.

Assume we then divide the sample of workers into two groups

- the first one includes those who did not receive training (say, group A)
- the second one includes those who did (group B)

We will then necessarily observe that the between groups (A and B) variance of earnings is significantly different from zero.

There are then two possibilities:

- either the within groups variance of earnings is important
- or it is not

In the latter case this would imply that the unobserved heterogeneity that was found to have a significant impact on the probability to receive training and on the earnings themselves is in fact the “hidden” criterion for labor market segmentation.

If however the within groups variance turns out to be important one would have to conclude that there is a great degree of overlapping between the two distribution of earnings, those of groups A and B.

It should then be clear that the division of the sample in two groups based on a distinction between those who received and those who did not receive on-the-job training is not relevant any more because the between groups variance turns out to be small compared to that of the within groups.

As a consequence on-the-job training (unless the unobserved heterogeneity has also an important effect on the within groups variance) cannot be in such a case a relevant criterion of labor market segmentation.

The main goal of this study is to show that new developments in income inequality decomposition techniques and in the application of such techniques to regression analysis (Fields, 2003) allow us to test the hypotheses mentioned previously.

We are in fact able to determine the exact impact of each variable

- not only on the overall variance of earnings
- but also on both the between and within groups dispersion (the groups referring here to those who received and did not receive on-the-job training).

Our study may thus shed new light on the link between training and earnings dispersion.

Outline

- first check the net effect (once the selectivity bias has been taken into account) of training on earnings
- second compare the relative importance of the between and within groups dispersions of earnings and hence find out whether there is a significant degree of overlapping between the distribution of earnings of the two groups previously mentioned
- quantify the exact contribution of the observed (the explanatory) variables and of the unobserved individual characteristics to the variance of earnings.

2. The data sources

A survey called «Continued Training 2000», a survey addressed to individuals (not firms), whether they received some kind of training or not.

The survey covered individuals who were less than 65 years old and had completed their initial formation

This survey is original because it covers all types of training, whatever their goal, whether they had a (direct or indirect) professional objective or whether their aim was more personal.

The types of training included were:

- Practical training: courses taken within the framework of continued training, seminars and conferences attended (training with a specialized “trainer” in a location different from the workplace)

- On the job training: this training takes place on the job with the help of a tutor and it implies the utilization of the usual work tools

- Self-formation: here the individual trains himself/herself, eventually with the help of specific tools (e.g. distance learning)

- Alternative periods of training: this type refers to work contracts that request a period of training, mainly what is called in France “qualification contracts”.

3. Estimating the Earnings Function

Some summary statistics

- on average individuals who underwent training earn 29.6% more than those who did not undergo any form of training
- the proportion of those who have a higher education diploma or a “baccalauréat” is higher among those who received a form of training
- the proportion of manual workers (whether “specialized” or “qualified”) is much higher among those who did not receive any training
- the proportion of those working full time is higher among those who received training

3.2 The Determinants of On-the-Job Training

We first estimated a Probit model which gives us the determinants of the access to the training that took place during the 14 months preceding the date at which the survey took place

The results of this Probit model are similar to those obtained in previous studies in France and show that

- the closer the links between the employer and the employee, the higher the probability of getting access to training. (Therefore individuals working under a contract of fixed duration are less likely to receive training than those employed under a contract of indeterminate duration).
- those who have the highest probability of receiving training are the “trainees” (“stagiaires” in French). These employees receive an intensive training in order to be integrated in the firm as quickly as possible.
- for full time contacts men are more likely than women to benefit from training but the opposite is true for part-time contracts (women working part-time seem to be employed in professions or sectors where the prevalence of training is usually important, see, Hanchane and Lambert, 2003)
- managerial staff and professions at the intermediate level of the hierarchy (“intermediate professions”) are the most likely to undertake a program of continued training
- firms with more than 500 workers are those that proportionally invest the most in training.

3.3 The Returns to Training and the Coefficients of the Earnings Function

The Probit model allowed us to estimate Mills' ratios (introduced in the regression to correct for selectivity bias).

In addition, following Barnow et al. (1980), we introduced a dummy variable equal to 1 for those who received training and to zero otherwise. The coefficient of this variable in the regression gives thus the impact on earnings of receiving training, after neutralizing the selectivity bias.

Other variables introduced in the regression:

- five dummy variables giving the educational level
- five dummy variables giving the qualification level of the job
- two dummy variables describing the type of contract
- age and its square
- gender and nationality
- seniority in the firm and its square
- three dummy variables giving the weekly duration of work
- two variables giving information on the type of work schedule

The results indicate that

- earnings grow with the level of education
- technicians, engineers and other members of the managerial staff earn more
- those having a work contract of *undetermined duration* earn more
- age and seniority have a non linear effect
- there is a selectivity bias (the coefficient of Mills´ ratio is significant)
- unobserved heterogeneity affects both the probability to receive on-the-job training and the earnings themselves. The “net” effect of on-the-job training on earnings is to increase the latter by 24% which is the difference between the values of the coefficients of the variable “on-the-job training” (0.50) and of Mills´ ratio (0.26).

Table 3: Regression Results

Dependent Variable: Logarithm of Monthly Earnings

Explanatory Variables	Coefficient	T-values
Constant	7.94	109.5
Higher Education Diploma	0.143	7.52
Holder of “Baccalauréat”	-0.064	-4.15
Holds a CAP or BEP	-0.164	-11.5
Holds a BEPC	-0.177	-9.31
Holds a CEP	-0.285	-18.6
“Specialized” Worker	-0.156	-6.96
“Qualified”Worker	-0.100	-5.65
Engineer or Managerial Position	0.356	21.1
Employee	-0.146	-9.41
Other Professions	-0.00034	-0.015
Has a Work Contract of Undetermined Duration	0.042	2.15
Other Categories of Trainees	-0.0062	-0.222
Seniority in Firm	0.011	8.03
Square of Seniority in firm	-0.00013	-3.26
Age	0.022	6.81
Square of Age	-0.00022	-5.54

Works Full Time (40 hours at least)	**0.666**	**49.8**
Works Part Time (30 to 40 Hours)	**0.406**	**19.8**
Works Part Time (less than 15 Hours)	**-0.814**	**-27.7**
Has the Same Work Schedule Every Day	**-0.041**	**-3.72**
Has a Variable Work Schedule	**-0.035**	**-2.72**
Female	**-0.148**	**-15.8**
Foreigner	**-0.062**	**-3.94**
Received Vocational Training	**0.479**	**9.91**
Mill's Ratio	**-0.249**	**-8.61**
R-Square	**0.670**	
Adjusted R-Square	**0.669**	
F-Value for Regression	**670**	

3.4 The role played by the side that took the initiative of the training or/and financed it

To better understand the various channels through which on-the-job training may have an impact on the earnings of those who received training Table 4 gives the results of a regression where we include only those individuals who received on-the-job training and the dependent variable is the residual of the earnings function

Table 4: Regression results

Dependent Variable: Residual of Regression of Table 2

Explanatory Variables	Coefficient of Regression	T-Values
Constant	-0.080	-2.2
Training originated in individual initiative*	0.107	2.7
Training originated in initiative from firm*	0.097	2.6
Training originated in initiative from both the individual and the firm*	0.100	2.6
The individual financed the training**	-0.104	-2.4
Other type of Financing**	-0.134	-5.6

* The reference category is “Other sources of initiative”

** The reference category is « Financing by the firm »

To better understand the complex links that may exist between the side taking the initiative of the training and that financing it, we made some additional tests. We checked in particular whether, among those who received a training that was financed by the employer, the cases where the initiative of such training was taken by the employer alone, had specific characteristics. It appears that as far as the field of training is concerned there is no real difference.

It seems that the individual is more involved when the duration of the training is longer and when it is validated by a diploma or a certification. This would seem to confirm that individuals look more for a form of training that is general rather than specific.

4. Estimating the Contribution of the Explanatory Variables to the Variance of Earnings: The Methodology

4.1. Estimating the Contribution of the Explanatory Variables to the Overall Variance (see, Fields, 2003):

We write the earnings function as

$y_j = \sum_{k=1 \text{ to } (K+3)} b_k z_{kj}$

y_j is the logarithm of the wage of the earnings of individual j

$z_{kj} = x_{kj}$ $\forall k=1$ to K, where x_{kj} refers to the value taken by the explanatory variable k for individual j. Note that these K variables do not include that referring to the participation (F_j) in the training program and the impact of the selectivity bias (λ_j). We therefore have also

$z_{K+1,j} = F_j$,

$z_{K+2,j} = \lambda_j$

$z_{K+3,j} = u_j$ where u_j is the value taken by the disturbance for individual j.

Note that we assume below that $b_{K+1} = c$, $b_{K+2} = d$, $b_{K+3} = 1$

Fields (2003) proved that

$$\sigma(y_j) = \sum_{k=1 \text{ to } (K+3)} [(b_k) \text{ Cor } (z_{kj}, y_j) (\sigma(z_{kj}))]$$

The relative contribution $s_k(y_j)$ of factor k to the dispersion $\sigma(y_j)$ is then expressed as

$$s_k(y_j) = [(b_k) \text{ Cor } (z_{kj}, y_j) (\sigma(z_{kj}))] / \sigma(y_j)$$
$$s_k(y_j) = [(b_k) \text{ Cov } (z_{kj}, y_j)] / V(y_j)$$

where $V(y_j)$ is the variance of the logarithms of earnings y_j.

So the relative contribution of factor x_h (h=1 to k) to this variance is

$s_h (y_j) = [(b_h) \text{ Cov } (z_{hj}, y_j)] / V(y_j)$

Similarly the relative contribution of the participation to an «on-the-job» training program may is

$s_F (y_j) = [(c) \text{ Cov } (F_j, y_j)] / V(y_j)$

The relative contribution of Mills' ratio is

$s_\lambda (y_j) = [(d) \text{ Cov } (\lambda_j, y_j)] / V(y_j)$

Finally the relative contribution of the unobserved variables (the disturbance u_j) is

$s_u (y_j) = \text{Cov } (u_j, y_j) / V(y_j)$

4.2. Contribution of the Explanatory Variables to the Within-Groups Variance

Introducing Mills' ratio we write for an individual belonging to group A (did not receive training) that

$y_{jA} = \sum_{k=1 \text{ to } K} b_k x_{kjA} + \rho\sigma_u [(- \phi_j)/(1- \Phi_j)] + w_{jA}$

where $\Phi(.)$ refers to the distribution function corresponding to the decision to participate in training and ρ is the correlation between the disturbances of the probit model and the earnings function, ϕ_j is the density function corresponding to Φ_j and σ_u is the standard deviation of the error term of the earnings function. Note that $[(- \phi_j)/(1- \Phi_j)]$ is the expression for Mill's ratio in group A and $\rho\sigma_u$ that for the coefficient d (see, Green, 2000).

For individuals in group B (received training) we write

$y_{jB} = \sum_{k=1 \text{ to } K} b_k x_{kjB} + c + \rho\sigma_u [\phi_j / \Phi_j] + w_{jB}$ (10)

where $[\phi_j / \Phi_j]$ is the expression for Mill's ratio in group B.

Since the within groups variance is equal to the weighted sum of the variance within each of the two groups A and B, the weights being the population shares (f and (1-f)) of the two groups, the contribution $s_{k,WITH(yj)}$ of each of the (K+2) factors to the within groups variance may then be written as

$$s_{k,WITH(yj)}=\{(1-f)[(b_k)Cov(z_{kj,j\in B}, y_{j,j\in B})/V_B(y_j)] + (f)[(b_k)Cov(z_{kj,\ j\in A}, y_{j,\ j\in A})/\ V_A(y_j)]\}$$

4.3 Contribution of the Explanatory Variables to the Between-Groups Variance

To compute the between-groups variance $V_{BET}(y_j)$ of the (logarithms of) earnings one has to neutralize the within groups dispersion and thus to assume that every worker who received on-the-job training receives the mean (logarithm of) earnings $y_{M,B}$ of those who received such training while those who did not receive any on-the-job are assumed to receive the mean earnings $y_{M,A}$ of those who did not receive any training.

The contribution $s_{k,B}(y_j)$ of each of the K first explanatory factors to the between groups variance, using again Fields' (2003) approach, will then be expressed as

$$s_{k,BET}(y_j) = [(b_k) \text{ Cov } (z_{kM}, yM)] / V_{BET}(y_j)$$

It is easy to show that

$$\text{Cov } (z_{kM}, y_M) = f (1-f) (x_{kMB} - x_{kMA}) (y_{MB} - y_{MA})$$

$$V_{BET} = f (1- f)(y_{MA} - y_{MB})2$$

So that

$$s_{k,BET}(y_j) = [(b_k) (x_{kMB} - x_{kMA})] / (y_{MB} - y_{MA})$$

For the contribution of the variable F_j to the between groups dispersion, one will obtain similarly, remembering that in this case $x_{kMB} = 1$ and $x_{kMA} = 0$,

$$s_{F,BET}(y_j) = [c / (y_{MB} – y_{MA})]$$

The contribution of the ratio of Mill λ_j to the between groups dispersion will be expressed as

$$s_{\lambda,BET}(y_j)=(\rho\sigma_\varepsilon)[(\phi_j/\Phi_j)_m-((-\phi_j)/(1-\Phi_j))_m]/(y_{MB} – y_{MA})$$

Finally the contribution of the disturbances to the between groups dispersion will be written as

$$s_{u,BET}(y_j) = [(w_{MB} – w_{MA})] / (y_{MB} – y_{MA})$$

where w_{MB} and w_{MA} are respectively the mean values of the disturbances in groups B and A.

We may therefore conclude, using all the previous results, that the contribution of a given factor k (k = 1 to K+3) to the total variance V_{TOT} of the logarithms of wages is the sum of three elements:

- its impact via its contribution to the within group A variance V_A
- its impact via its contribution to the within group B variance V_B
- its impact via the between groups variance V_{BET}

The exact formulations are given in the paper.

5. Decomposing the Variance of Earnings: The Results

Table 6 gives the decomposition of the total variance of the logarithm of wages into two components, the between groups (the groups being those who received training and those who did not) and the within groups variance. It appears that most of the dispersion (94.5% of the variance) takes place within groups while the between groups variance represents only 5.5% of the total variance.

Table 6: Breakdown of the Total Variance into the Sum of Between and Within Groups Variances

Type of Variance	Value and Share (in percentage)
Overall Variance	32.33 (100%)
Between Groups Variance	1.78 (5.5%)
Within Groups Variance	30.55 (94.5%)

5.1 Contributions of the various variables to the within groups variance:

To analyze the results we have to remember that the contribution of a given variable k (in percentage) to the variance of the regression is a function

- of the coefficient b_k of this variable in the regression (earnings function)
- its correlation Cor (z_{kj} , y_j) with the dependent variable (the logarithm of earnings)
- its dispersion, relative to the dispersion of the dependent variable $\sigma(z_{kj})$ / $\sigma(y_j)$.

Such a general interpretation evidently holds also in the case of the within groups variance.

Table 8 indicates that the three variables that contribute most to the dispersion of (the logarithms of) earnings in group A are respectively

- the dichotomous variables “works full-time” and “works part time, 0 to 15 hours”
- the residuals

Since the dependent variable represents the monthly earnings the role played by the number of hours of work is not surprising.

Once this factor is taken into account, it thus appears that approximately half of the remaining variance is explained by the unobserved heterogeneity of the individuals.

The two other variables that have a somehow significant contribution to the variance of the (logarithms of) earnings in group A are the dichotomous variables “holds a CEP” and “employee”.

In group B (those individuals that received training) the two most important contributions to the variance of earnings are again those of the residuals and of the variable “worked full time”.

The explanation is similar to that given earlier for group A. The contribution (in percentage) of these two variables is however smaller because for the group of individuals who received training other variables play a role. First note the important contribution of the dummy variable “Engineer or Managerial Position”. Second the other variable whose contribution should be mentioned is Mill’s ratio (see below).

Table 8: Contributions of the Explanatory Variables to the Within Groups Variance

Explanatory Wariables	Individuals Belonging to Group A (Did not Receive any Vocational Training)	Individuals Belonging to Group B (Received Vocational Training)
Higher Education Diploma	1.31	3.07
Holder of “Baccalauréat”	-0.21	0.065
Holds a CAP or BEP	-0.30	1.96
Holds a BEPC	0.18	0.49
Holds a CEP	5.90	3.36
“Specialized” Worker	1.58	0.88
“Qualified” Worker	-0.75	1.03
Engineer or Managerial Position	6.91	14.67
Employee	4.13	5.12
Other Professions	-0.0002	0.0002
Has a Work Contract of Undetermined Duration	0.20	0.366
Other Categories of Trainees	0.0073	0.026
Seniority in Firm	4.29	5.44
Square of Seniority in firm	-1.60	-1.86
Age	-1.24	-7.78
Square of Age	2.01	10.1
Works Full-time (40 hours at least)	25.97	18.19

Works Part-time (30 to 40 hours)	**-1.83**	**-1.59**
Works Part-time (less than 15 Hours)	**9.40**	**2.66**
Has the Same Work Schedule Every Day	**0.083**	**0.060**
Has a Variable Work Schedule	**0.055**	**-0.15**
Female	**5.00**	**4.29**
French	**0.15**	**0.008**
Mill's Ratio	**3.52**	**7.72**
Residuals	**35.49**	**33.01**

5.2 Contributions of the various variables to the between groups variance

We have mentioned previously that the contribution of a given variable to the between groups variance is

- positively related to the coefficient of this variable in the regression and to the difference between the two groups in the means of this variable
- negatively related to the difference between the two groups in the means of the dependent variable (logarithms of earnings).

Four of the five variables that have an important impact on this between groups variance were already mentioned when we analyzed the determinants of the within groups variance. These variables are by decreasing order of importance

- “Holds a CEP” (17%)
- “Works Full Time” (15%)
- “Is an Engineer or Has a Managerial Position” (13%)
- “Is a Specialized Worker” (6.1%)
- “Works less than 15 hours per week” (5.1%).

The impact of Mills ratio and of the variable “received vocational training” will be discussed below in a separate section.

Table 7: Contributions of the Explanatory Variables to the Between Groups Variance

Explanatory Variables	Contribution (in percentage)
Higher Education Diploma	3.55
Holder of “Baccalauréat”	-1.31
Holds a CAP or BEP	1.82
Holds a BEPC	0.35
Holds a CEP	17.04
“Specialized” Worker	6.10
“Qualified”Worker	2.71
Engineer or Managerial Position	13.13
Employee	2.14
Other Professions	-0.28
Has a Work Contract of Undetermined Duration	0.32
Other Categories of Trainees	0.003
Seniority in Firm	1.53
Square of Seniority in firm	-0.37
Age	9.61
Square of Age	-11
Works Full-time (40 hours at least)	15

Works Part-time (30 to 40 hours)	**-0.61**
Works Part-time (less than 15 hours)	**5.07**
Has the Same Work Schedule Every Day	**0.75**
Has a Variable Work Schedule	**-0.12**
Received Vocational Training	**161.8**
Mill's Ratio	**-127.6**
Residuals	**0**

5.3 The contributions of Mills´ ratio and of the variable “Received Vocational Training” to the variance of earnings

Let us fist look at the contribution of Mills’ ratio to the within groups variance of earnings. It appears (see, Table 8) that 3% of the variance of earnings among those who did not receive any training (group A) is due to differences among the individuals in the value taken by the Mills ratio, that is to that part of the unobserved heterogeneity that has an impact on the a priori probability to receive such a training. Among those who did in fact receive such a vocational training, the contribution of Mills ratio to the variance of earnings is even equal to 8%.

As far as the between groups variance of earnings is concerned, we have to take into account the contributions of the dummy variable “Received Vocational Training” a well as that of the Ratio of Mills. The combined contribution of these two variables to the between groups variance of earnings may be considered as the net contribution of training to the between groups variance of earnings. Such a contribution takes into account not only the impact of training on earnings but also the fact that the unobserved heterogeneity has an effect on the probability of receiving training. It thus appears that 35% (170 – 135) of the between groups variance of earnings is due to this combined effect. Note that the sign of the contribution of the Mills ratio is negative.

Clearly the allocation of workers to training is not random since the unobserved heterogeneity has an impact on the probability to receive training as well as on the earnings themselves.

Let us now summarize the breakdown.

Table 10: Breakdown of the Overall Variance of the Logarithm of Wages

Type of Variance	Value
Between Groups Variance of Actual (Logarithms of) Incomes	0.018
- Contribution of Vocational Training (Dummy Variable "Received Vocational Training" Plus Mills Ratio)	0.006
- Contribution of Other Variables	0.012
Variance of Predicted (Logarithms of) Incomes	0.213
- Between Groups Variance of Predicted (Logarithms of) Incomes	0.018
- Within Groups Variance of Predicted (Logarithms of) Incomes	0.195
- Contribution of the group who did not receive any vocational training (group A)	0.145
- Contribution of the group who received vocational training (group B)	0.050
Variance of Actual (Logarithms of) Incomes	0.323
- Between Groups Variance of Actual (Logarithms of) Incomes	0.018
- Within Groups Variance of Actual (Logarithms of) Incomes	0.306
- Contribution of the group who did not receive any vocational training (group A)	0.230
- Contribution of the group who received vocational training (group B)	0.076

6 Conclusions

This paper attempted to devise a methodology that allows estimating the exact impact of training on the dispersion of wages, extending an approach originally proposed by Fields (2003).

The empirical illustration is based on a survey conducted in France at the end of the twentieth century.

The results of the analysis show first that when a distinction is made between workers who received training and those who did not, the between groups dispersion explains only 5.5% of the overall variance of earnings.

We also found that more than one third of this between groups variance was explained by the combined effect of the unobserved heterogeneity and the distinction between those who received and did not receive on-the-job training.

We also noted that the unobserved heterogeneity led to a drastic reduction of the impact of training on earnings, since those who received training were also those who had a priori the highest probability of receiving training.

Most of the earnings dispersion however turned out to be a within groups dispersion and more than two thirds of this within groups variance of the logarithms of earnings could be explained by the variables that were taken into account.

Given that there is a small between groups and a big within groups dispersion, there is a lot of overlapping between the distributions of earnings of the two groups, those who received and those who did not receive training.

Such findings imply indeed that the unobserved heterogeneity plays a key role in the selection of those who receive training and thus indirectly has an impact on the difference between the *average* earnings of those who receive and do not receive training. It cannot however be considered as a variable that could lie behind market segmentation.

This is so because the within groups variance is much higher than that of the between groups so that the distributions of earnings of these two groups show a great degree of overlapping. In other words there is a much greater degree of heterogeneity within than between the two groups corresponding to those who received and did not receive on-the-job training. As a consequence we believe that if labor market segmentation exists, it must be based on other criteria.