Training, Wage Growth, and Job Performance: Evidence from a Company Database

Ann P. Bartel, *Columbia University and National Bureau of Economic Research*

A unique dataset collected from the personnel records of a large company is used to study the relationship between on-the-job training and worker productivity. The analysis shows how information contained in a company database is useful for eliminating heterogeneity bias in the estimation of training's impact on wages and job performance. Even when selection bias in assignment to training programs is eliminated, training is found to have a positive and significant effect on both wage growth and the change in job performance scores, thereby confirming the robustness of the relationship between training and productivity.

I. Introduction

According to the theory of human capital, investments in training lead to increases in worker productivity. In order to test the validity of this prediction directly, researchers require data on investments in training and on-the-job productivity. In the absence of such data, however, early research on this topic used information about the shape of wage profiles to make inferences about the relationship between human capital investments and productivity (Mincer 1974). The problem with this approach was that it

An earlier version of this article was presented at the National Bureau of Economic Research (NBER) Summer Institute Workshop on the Economics of Labor, July 1991. Comments from the participants in the NBER Workshop and the Columbia University Labor Economics Seminar are greatly appreciated.

[*Journal of Labor Economics,* 1995, vol. 13, no. 3]

0734-306X/95/1303-0003$01.50

was difficult to refute alternative theories that imply rising wage profiles over the life cycle but have nothing to do with investments in training.[1]

More recently, the availability of data on training has allowed researchers to analyze directly the link between on-the-job training and the pattern of wages (e.g., Lillard and Tan 1986; Barron, Black, and Loewenstein 1989; Brown 1989; Holzer 1990; Mincer 1991; Lynch 1992; Bartel and Sicherman 1994). With the exception of the work by Barron, Black, and Loewenstein and Holzer, all of these papers used data on training that were reported by the individual employee. There are a number of problems with self-reported training information. For example, individuals may have difficulty recalling all training events that occurred during the past year, especially the duration of such events. In a recent study of matched employer and employee responses, Barron, Berger, and Black (1994) show that employees and employers differ in their reports of the amounts of both formal and informal training, with workers reporting less training. Another problem is the comparability of training events across individuals working in different organizations; that is, is a day of training in company A the same as a day of training in company B?

In addition to these generic problems with self-reported information, there are problems associated with specific data sets. For example, the papers that used the Panel Study of Income Dynamics suffer from having to use information on how long it took the "average" person to become qualified for the job, not how long the respondent actually took to become qualified.[2] The papers by Bartel and Sicherman and Lynch used the National Longitudinal Survey (NLS) Youth Cohort, which contains very comprehensive data on private sector training. The weakness of that data set is that, for the time period 1979 through 1987, training duration data are either incomplete or inaccurate, and even the post-1987 surveys do not accurately measure duration of very short training programs.[3]

[1] For example, Lazear (1981) discusses how firms offer upward-sloping wage profiles to their workers in order to discourage shirking. Salop and Salop (1976) suggest that upward-sloping profiles are used by firms as a way of discouraging "movers" from applying for jobs. Jovanovic (1979) demonstrated how job-matching under imperfect information could generate upward-sloping wage profiles. Finally, Loewenstein and Sicherman (1991) showed that increasing wage profiles could reflect workers' preferences.

[2] For more details on this measure of training, see Sicherman (1991).

[3] In the National Longitudinal Survey of Youth (NLSY) surveys that were conducted prior to 1988, information on duration is not reported if the training spell lasted less than 4 weeks. For spells that lasted more than 4 weeks, duration is measured inaccurately because the respondent reports only the starting and ending months of the training program. Using the starting and ending months to measure elapsed duration seriously overestimates true duration if, for example, the individual trained only 1 or 2 days per week over the interval. Starting with the 1988 NLSY survey, duration of training is reported in terms of weeks for all spells. Training spells that last between 1 and 5 days are all coded as 0 weeks.

The papers by Barron, Black, and Loewenstein (1989) and by Holzer are unique in that they use the (EOPP) data set in which data on training were collected from employers and include information on both formal and informal training. The main disadvantage of the EOPP data, however, is that the information on training is reported only for the most recent hire in the firm, thus limiting the applicability of the results to longer tenure workers.

In this article, I study the relationships between formal on-the-job training and wages and performance by using the personnel records of a large manufacturing firm. There are many advantages to using a data set collected from one firm. First, the problem relating to individuals' recalling the type and amount of training they received is avoided since all training spells and their exact durations are recorded on the individual's personnel record. Precise measures of training duration, even for programs that last only a few days, are obtained. Second, since the individuals are all trained by the same firm, there is no bias resulting from definitions of training varying across diverse firms. Third, unlike the EOPP data set, the training experiences of all employees, not just the most recent hire, are recorded. Data of this nature provide the researcher with a unique opportunity to compare the careers of individuals in the same firm and analyze how wage growth within the firm is influenced by training received in that firm. Although these advantages are important, it must be recognized that the findings obtained with a company database are not necessarily generalizable to other companies, especially to those in different industries. In addition, my analysis refers only to the professional employees in the company, most of whom are scientific, information systems, or engineering staff, and the results may not carry over to other employee groups.

On balance, however, using a company database makes an important contribution to the literature on empirical testing of human capital theory. In addition, the particular database that I use contains unique information that enables me to provide a more thorough analysis of the relationship between training and worker productivity. First, the database includes performance ratings which provide the unusual opportunity of using two alternative measures of the worker's productivity in the firm, that is, both wages and performance scores. The impact of training on the individual's performance score can be compared to training's effect on wages, in order to test the robustness of the measured relationship between training and productivity on the job.[4] Second, the database contains information on the relationship between an employee's salary and the salaries of other employees in his job category. I show how this information can be used to

[4] Previous work by Medoff and Abraham (1981) used a company database to provide a direct test of the relationship between company experience and performance.

Table 1
Summary Statistics for 1990

	Mean	Standard Deviation
Years of schooling	16.53	2.03
Previous experience	7.01	6.65
Company experience	7.41	8.35
Age	35.91	10.09
Monthly salary ($)	3,972.83	949.98
Occupational distribution:		
Finance	.048	
Engineering	.141	
Manufacturing	.090	
Sales	.078	
Information systems	.154	
Scientific	.440	
Staff	.030	

correct for unobserved heterogeneity bias that can plague single equation estimates of the impact of training on wage growth.

In the next part of the article, the data set is described. The econometric framework is discussed in Part III, and results are presented in Part IV. Part V summarizes the analysis and discusses implications for future research.

II. Data

Data for this study are taken from the 1986–90 personnel records of a large manufacturing company. Individuals who were classified by the company as professional employees were selected for analysis, resulting in a total sample of 19,000 observations, averaging 3,800 individuals per year. These employees are distributed across eight functional areas in the company: (1) finance, (2) engineering, (3) manufacturing, (4) marketing, (5) information systems, (6) R&D, (7) staff services, and (8) support services. The types of occupations held by these individuals include accountants, engineers, purchasing agents, quality control planners, market researchers, systems analysts, bench scientists, human resource professionals, and industrial hygienists. For each individual who appears in the company's database in a particular year, information is reported on length of service, source of hire, years of education, salary, tenure on the current job in the company, and days spent in formal training in the past year. Information on performance ratings is available for the years 1989 and 1990 only.

Table 1 contains summary statistics for the professional employees in the firm. This group of individuals is highly educated and well paid. In 1990, the mean years of schooling was 16.45 and the average monthly salary was $3,700. The average professional employee was 36 years old, worked for 7 years before coming to this company, and had approximately 8.3 years of experience at the company.

Table 2
Representative Course Titles in the Three Training Categories

	Length of Course (Days)
"Core" training program:	
1. Management for Managers	5
2. Management for Middle Managers	5
3. Managing and Coaching Performance	2
4. Leadership Thinking	3
Corporate employee development program:	
1. Oral Presentations Workshop	2
2. Effective Written Communications	3
3. Problem Solving and Decision	4
4. Managing Performance	2
5. Performance Appraisal and Salary Administration	2
6. Stress Management	1
7. Time Management	2
Technical programs:	
1. Project Management	2
2. Clinical Statistics	3
3. Quality Control	6
4. Good Manufacturing Practices	1
5. Introduction to Computer Programming	5

The company has a fairly substantial training budget. In 1990, it spent approximately $1,950 per employee on formal training. This can be compared to the average figure for all U.S. firms, which was approximately $385 per employee in 1989.[5] The company offers a wide range of training programs for its professional employees. Table 2 provides a listing of representative course titles in each of three groups of training programs. One group of training programs is called the "Core Program." Courses in this program are designed for any individual in the company whose job involves supervising at least one other employee. These courses teach the individual how to evaluate and improve employee performance, how to effectively manage time, how to be an effective leader, and how to implement change. A second group of programs is called "Corporate Employee Development." The courses in this category involve learning such skills as problem solving, decision making, written and oral communication, improvement of job performance, and time management. Third, the company offers courses in computer skills, information systems, research skills, good manufacturing practices, job safety, and other technical areas. All training programs are offered on a full-day or half-day basis and, as table 2 shows, typically last between 2 and 5 days.

Table 3 provides descriptive statistics on the training received by the professional employees in the firm. For each year, the percentage of employees receiving training and the mean days in training for those receiving

[5] See U.S. Congress, Office of Technology Assessment (1990).

Table 3
Percentage Receiving Training and Mean Days in Various Programs

	All Programs	"Core" Program	Employee Development Program	Technical Programs
1986 (N = 3,118):				
Percentage receiving training	55.4	0	24.4	39.1
Mean days for those in programs	3.4	0	2.2	3.4
1987 (N = 3,356):				
Percentage receiving training	48.0	.1	28.7	28.7
Mean days for those in programs	3.4	4.0	2.6	3.0
1988 (N = 3,690):				
Percentage receiving training	56.2	1.9	36.1	33.2
Mean days for those in programs	3.3	4.2	2.9	2.2
1989 (N = 4,091):				
Percentage receiving training	63.7	3.0	35.4	45.4
Mean days for those in programs	4.4	4.0	2.8	3.7
1990 (N = 4,508):				
Percentage receiving training	57.9	3.5	28.7	41.5
Mean days for those in programs	3.5	4.2	2.5	2.7

training are shown for three categories of training: (1) "core" training, (2) employee development, and (3) technical programs. The last category covers all courses that are not listed in either the core program or the employee development program; for professional employees, this category largely covers research, computer, and technical skills. The data in table 3 show that, during each of the 5 years for which data are available, at least half of the professional employees received some formal training.[6] Very little of this training was in the "core" training program because the job of a professional employee often does not involve supervisory activities.[7] The mean days spent in training per year for those professional employees who received some formal training during the year ranged from 3.3 days to 4.4 days. Since these figures are remarkably close to the numbers reported

[6] In 1989, the percentage receiving formal training (63.7%) is higher than the corresponding percentages for the other years. In that year, the company introduced several new technical training programs and participation in technical training increased compared to previous years.

[7] Since the "core" program was introduced at the end of 1986, there were no participants in 1986.

in the annual survey conducted by *Training Magazine,* we can reasonably conclude that the company under study is typical of the companies included in the magazine's nationwide survey.[8]

Since training is more likely to occur during the initial stages of an individual's tenure, the data in table 3 need to control for length of service at the company. This is done in table 4, where the probability of receiving training and the days spent in training are calculated separately for each year of hire. In this table, the three training categories are aggregated. Reading across a row in table 4 shows the "cross-sectional" effect of tenure on the amount of training. In all years, the probability of receiving training and the number of training days conditional on receipt of training peak in the second year for which an entry in the table appears, for example, 72.0% and 4.3 days for training received in 1990 by individuals hired in 1989. The reason for this is as follows. Individuals are hired during all months of the year. If a hire took place during the latter half of the year, for example, October 1987, training in the first "year" on the job will not appear until the second calendar year, in this case 1988. While this phenomenon would cause some of the early training an employee receives to be recorded in the second year, by itself it should not produce a peak in the second year. The second year peaking could occur if the company uses a probation period, that is, training does not begin until after the accumulation of several months of tenure.

Although the probability of receiving training and the amount of training are highest for the newly hired employees, an important observation is that experienced employees at this company continue to receive formal training. For example, in 1990, 47% of the individuals who were hired before 1980 received some formal training. This is a significant finding since it indicates that training is not confined to new hires but is an ongoing process at this company. In the next section, a model of the incidence of training is developed that takes account of the importance of training beyond the first year of tenure.

The data in table 4 can also be used to follow a given cohort over time. Starting from the bottom of a column and reading up shows the trend

[8] Each May, the publishers of *Training Magazine* conduct a survey of American businesses on the amount of time their employees spend in formal training during the year. The surveys for 1989 and 1990 showed that, for individuals who received some formal training, the average amount of time spent in formal training was approximately 4 days. Data on time spent in formal training in surveys of low-wage workers with high school educations show that these individuals spend more time in formal training. For example, Holzer (1990) shows that, in the EOPP Survey, 10.9 hours was spent by specially trained personnel in providing the most recently hired worker with formal training during the first 3 months of employment. Assuming that training proceeded at the same rate during the year for those individuals, that would translate into 44 hours or 5.5 days of formal training.

Table 4
Percentage Receiving Training and Mean Days for Those Who Received Training

	Year of Hire												
	All	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	Pre-1980
1. 1990:													
% trained	57.9	55.8	72.0	68.3	65.7	63.1	53.3	52.6	58.9	53.1	50.8	52.0	47.0
Mean days	3.5	3.5	4.3	3.4	3.4	3.2	3.4	2.7	2.7	3.7	3.1	3.4	3.1
2. 1989:													
% trained	63.7	...	60.1	76.6	68.9	67.8	64.8	67.4	64.7	63.2	62.2	60.7	57.1
Mean days	4.4	...	3.8	4.8	4.8	4.4	4.5	4.2	4.5	4.7	5.3	5.3	4.0
3. 1988:													
% trained	56.2	...	...	59.4	67.8	60.8	63.4	61.6	59.3	54.5	57.3	56.1	47.7
Mean days	3.3	...	...	2.9	3.9	3.1	3.1	3.2	3.6	3.5	3.7	3.9	3.0
4. 1987:													
% trained	48.0	...	...	...	46.1	62.2	59.0	53.7	54.6	44.5	55.3	50.9	41.2
Mean days	3.4	...	...	...	3.0	4.0	3.7	3.5	3.8	3.2	3.6	3.6	3.1
5. 1986:													
% trained	55.4	...	...	...	...	42.9	60.2	56.5	59.1	60.1	59.6	57.7	55.1
Mean days	3.4	...	...	...	...	3.0	3.8	3.4	3.5	4.2	3.3	3.6	3.2

over time in the amount of training received by a group of individuals hired during a given year. Again we observe that, while training incidence tends to peak during the employee's second year at the firm, training persists throughout the employee's career.

III. Econometric Framework

At time period t, an individual's productivity, $\Pi_{i,t}$, is a function of the stock of human capital he has accumulated as a result of investments made prior to and including the present period:

$$\Pi_{i,t} = \mathrm{f}\left(\sum_{n=1}^{t} \mathrm{INV}_{i,n}\right), \tag{1}$$

where $\mathrm{INV}_{i,n}$ is the individual's investment in human capital in time period n. Empirical estimates of equation (1) typically involve using the individual's wage rate as a proxy for productivity, and length of schooling, labor market experience, and tenure with the current employer as proxies for investments in human capital.

Access to a company database enables the researcher to use more appropriate measures of the variables in equation (1). The information that I have on days of formal training in the company can be used in place of, or in addition to, the more commonly used proxy, company tenure. For example, an earnings equation of the following form could be estimated:

$$\ln(\mathrm{SAL}_{i,t}) = \alpha + \beta_1 X_{i,t} + \beta_2 \mathrm{STKTRAIN}_{i,t} + \beta_3 \mathrm{YEAR} + \mu_i + \nu_{i,t}, \tag{2}$$

where the vector X might include such variables as years of education, labor market experience prior to joining the company, the square of previous labor market experience, length of service at the company, the square of length of service, and occupation. What makes equation (2) different from the standard earnings equation is the variable $\mathrm{STKTRAIN}_{i,t}$, which measures the stock of formal training, that is, the total days of formal training that the individual has accumulated in the company. Since $\mathrm{STKTRAIN}_{i,t}$ measures formal training only, length of service needs to be included in equation (2) in order to capture investments in informal training. Note that the error term in equation (2) is composed of a person-specific fixed effect, μ_i, and a person-specific effect that varies over time, $\nu_{i,t}$.

While equation (2) is superior to the equations typically estimated on such databases as the NLS and the Panel Study of Income Dynamics (PSID), it still has two weaknesses. The first is that the company database contains no information on training received prior to 1986, and therefore, STKTRAIN can be accurately measured only for individuals hired after 1985. Another problem with equation (2) is that some of the regressors may be correlated with the error term because of the role of person-specific effects. If, for example, more able or more promising employees receive more training, then the coefficient on STKTRAIN will, at least in part, reflect the role of these unobserved characteristics. As with other panel data sets, this problem can be addressed here by estimating a first-difference version of equation (2). In particular, annual changes in salary can be regressed on changes in the right-hand variables. In this formulation, all time-invariant effects (both observed and unobserved) drop out of the equation, leaving only time-varying variables. This results in equation (3):

$$\begin{aligned} \ln(\mathrm{SAL}_{i,t}) - \ln(\mathrm{SAL}_{i,t-1}) &= \beta_1(X_{i,t} - X_{i,t-1}) \\ &+ \beta_2(\mathrm{STKTRAIN}_{i,t} - \mathrm{STKTRAIN}_{i,t-1}) + \nu_{i,t} - \nu_{i,t-1}. \end{aligned} \tag{3}$$

In addition to eliminating omitted variable bias due to unobserved person effects that are constant over time, equation (3) has the advantage of being able to be estimated on the complete sample, not just individuals hired after 1985.

Although the specification in equation (3) is an improvement over equation (2), the estimated coefficient on the training variable, β_2, may still be biased if the new error term, $\nu_{i,t} - \nu_{i,t-1}$, is correlated with the probability of receiving company training. The firm may, through observing the employee's behavior, gather information about his true capabilities and use this information to select an employee for training in order to facilitate his within-firm career advancement. While the correction embedded in equation (3) eliminates the bias resulting from a fixed person effect, it does not eliminate the bias that results from time-varying heterogeneity. Assignment to employee training programs is more likely to be influenced by time-varying heterogeneity than fixed person effects, as individuals, through their performance, reveal their potential for advancement. This type of bias can be eliminated by modeling the determinants of training inside the firm and obtaining an instrument for $(\mathrm{STKTRAIN}_{i,t} - \mathrm{STKTRAIN}_{i,t-1})$ that is uncorrelated with the error term in equation (3).

A training incidence equation will be estimated in which the probability of receiving training during the interval of time between the last day (i.e., December 31) of year $t - 1$ (when salary in year $t - 1$ is recorded) and the last day of year t (when salary in year t is recorded) is the dependent variable.[9] The company database includes a unique variable that can be used in the training equation to identify an instrument for $(\mathrm{STKTRAIN}_{i,t} - \mathrm{STKTRAIN}_{i,t-1})$ that is likely to be uncorrelated with $(\nu_{i,t} - \nu_{i,t-1})$. In

[9] Alternatively, the duration of training that occurred during this time interval can be used as the dependent variable.

particular, in addition to the data on the individual's own salary, the company records, at the end of each year, the employee's relative salary, that is, how his or her salary compares to the salaries of other individuals performing the same job. This information is expressed in ratio form; that is, an individual whose salary is 90% of the salaries of other employees performing the same job would have the value .90 for this relative status measure. The identities of the employees whose salaries reflect the denominator of this ratio are not revealed; only the ratio is reported. For the empirical analysis, this ratio is denoted $\text{RELSTAT}_{i,t-1}$.

It can be argued that an individual's relative status in his job as of the end of time period $t - 1$ provides a signal to the company as to the likely payoff from training the individual during time period t. Consider the following example. As of the end of time period $t - 1$, two individuals have identical years of service with the company and the same tenure in a particular job, but one individual has a higher salary relative to other individuals performing this job. If training is nonremedial, but rather a foundation for career advancement, the individual with high relative status should be more likely to be selected for training. If training is remedial, however, the opposite prediction would be made: an individual with lower relative status would be selected for training in order to bring his skills up to the "average" level.

There are a number of reasons, however, why this approach needs to be modified for individuals who started a new assignment in the company during time period $t - 1$. First, this individual may, as a result of his new assignment, be required to take certain training courses. Second, in my discussions with company executives I was advised that $\text{RELSTAT}_{i,t-1}$ is not a meaningful indicator of relative status for individuals whose tenure in the particular job is less than 1 year. These considerations lead to the following specification of the training equation:

$$\begin{aligned}\Delta\text{STKTRAIN}_{i,(t,t-1)} = \alpha_0 &+ \alpha_1 Z_{i,t-1} \\ &+ \alpha_2\text{RELSTAT}_{i,t-1}(1 - \text{NEWJOB}_{i,t-1}) \\ &+ \alpha_3\text{NEWJOB}_{i,t-1} + \alpha_4\text{YEAR} + \eta_{i,t}.\end{aligned} \tag{4}$$

The dependent variable in equation (4) is the incidence or amount of training received between the end of year $t - 1$ and the end of year t, that is, during the 12 months of year t. The vector $Z_{i,t-1}$, measured at the end of year $t - 1$, includes years of service on the job held as of the end of time period $t - 1$, years of education, years of company service, source of hire (i.e., whether hired as a result of campus recruiting, advertisement, employment agency, or employee referral), and type of occupation in the company. Source of hire is included in this equation as a control for the prehire skill endowment. The variable $\mathrm{RELSTAT}_{i,t-1}$ is the individual's relative salary measured at the end of year $t - 1$, and $\mathrm{NEWJOB}_{i,t-1}$ is a dummy variable that equals 1 if the job the individual held at the end of year $t - 1$ began during year $t - 1$.

According to the specification in equation (4), $\mathrm{RELSTAT}_{i,t-1}$ is used by the firm to select individuals for training if they have been in their jobs at least 1 year. Recall from table 4 that, at this company, training beyond the first year of tenure is quite common. If training is nonremedial, α_2 should be positive; if training is remedial, it will be negative. In addition, we would expect α_3 to be positive if individuals who are new to their jobs are more likely to receive company training. The estimates from equation (4) will be used to construct an instrument for $\Delta\mathrm{STKTRAIN}_i$ which will then be used in equation (3) to obtain an unbiased estimate of the effect of training on salary growth.[10] All standard errors in equation (3) will be corrected utilizing the procedure suggested by Murphy and Topel (1985) for two-step econometric models.[11]

One problem with this approach is that it might be argued that $\mathrm{RELSTAT}_{i,t-1}$ belongs in both the wage growth equation and the training equation. Employees with high values of $\mathrm{RELSTAT}_{i,t-1}$ may be "stars" who are selected for training programs but who also receive large wage increases irrespective of the training that they receive. If this is an accurate description of career paths at the company, $\mathrm{RELSTAT}_{i,t-1}$ should have a positive and significant effect on wage growth, even if the predicted value of training is included in the wage growth equation. It is possible that the inclusion of $\mathrm{RELSTAT}_{i,t-1}$ in the wage growth equation may reduce the coefficient on training to insignificance if the true determinant of career advancement in the company is $\mathrm{RELSTAT}_{i,t-1}$, not the acquisition of training. Even if the latter is true, however, $\mathrm{RELSTAT}_{i,t-1}$ may still be used by the firm to select employees for training, and we will observe a positive and significant coefficient on $\mathrm{RELSTAT}_{i,t-1}$ in the training equation, equation (4). In order to test the alternative hypotheses, equation (3) will be estimated with and without $\mathrm{RELSTAT}_{i,t-1}$.[12]

[10] The salary growth equation can be estimated only on the sample of individuals who are in the firm in both time periods $t - 1$ and t. Company documents indicate that individuals who leave the firm are those who tend to have low performance ratings compared to their peers. The effect of this type of turnover is to reduce heterogeneity bias since the "ability" distribution of the remaining employees will be less dispersed.

[11] Alternatively, equations (3) and (4) can be estimated using two-stage least squares in which equation (4) is treated as a linear probability model. Since I am interested in reporting the logit estimates for equation (4), I prefer to use the two-step estimation method described in the text.

[12] When $\mathrm{RELSTAT}_{t-1}$ is included in the wage growth equation, the source of hire variables and NEWJOB_{t-1} can identify the instrument for training.

In addition to data on salaries, the company database includes information on performance ratings recorded in January 1989 and January 1990. This provides a unique opportunity to measure the effect of training on the individual's productivity growth as proxied by the change in performance ratings. Referring back to equation (1), the individual's performance score can serve as an alternative measure of Π, the individual's productivity. Equation (3) can be respecified with the dependent variable as the change in performance ratings:

$$\begin{aligned} \mathrm{PERFRAT}_{i,t} - \mathrm{PERFRAT}_{i,t-1} &= \beta_1(X_{i,t} - X_{i,t-1}) \\ + \beta_2(\mathrm{STKTRAIN}_{i,t} - \mathrm{STKTRAIN}_{i,t-1}) &+ \nu_{i,t} - \nu_{i,t-1}, \end{aligned} \tag{5}$$

where $\mathrm{PERFRAT}_{i,t}$ is the individual's performance score in January 1990 and $\mathrm{PERFRAT}_{i,t-1}$ is his score in January 1989. Unfortunately, performance rating scores cannot be compared across jobs, and, therefore, equation (5) can be estimated only for the sample of individuals who did not change jobs during the relevant time period.[13] The instrument for $\Delta\mathrm{STKTRAIN}_i$ obtained from equation (4) will be used in equation (5) to obtain an unbiased estimate of the effect of training on the change in performance scores. As in the case of equation (3), all standard errors in equation (5) will be corrected using the Murphy and Topel (1985) procedure. Estimates of equation (5) will measure the direct impact of training on the change in job performance and can be contrasted with the estimates from equation (3) of the effects of training on salary growth. Since performance scores are measured as discrete variables, the actual estimation of equation (5) will use a multinomial logit model on the determinants of going up, going down, or staying the same in the measure of performance. This will be described in more detail in the next section.

In summary, the information in the company database will be used first to obtain estimates of equation (4), the determinants of training. Predicted values for investment in training, $(\mathrm{STKTRAIN}_{i,t} - \mathrm{STKTRAIN}_{i,t-1})$, will be calculated from these estimates and then used in the estimation of equation (3) to obtain an unbiased estimate of the impact of company training on wage growth. Equation (5) will be estimated using the predicted value for $(\mathrm{STKTRAIN}_{i,t} - \mathrm{STKTRAIN}_{i,t-1})$ to obtain an unbiased estimate of the impact of company training on the change in performance scores.

IV. Empirical Results

In this section, I present the results of estimating equations (3), (4), and (5) described above. Results for the determinants of receiving training are

[13] This restriction is described in more detail in the next section of the article.

discussed first, followed by the analysis of the effects of training on salary growth and performance scores.

A. The Determinants of Training

Table 5 shows the results of estimating several variants of equation (4), the change in the individual's training stock between the end of time period $t - 1$ and the end of time period t. In the first four columns, the dependent variable is the probability of receiving training during the time period, and the equations are estimated using binomial logit. In columns 5–8, the dependent variable is the number of days of training the individual received, and the equations are estimated using the maximum likelihood tobit procedure.[14] Columns 1 and 5 include all three types of training, that is, in column 1, the dependent variable equals 1 if the individual received any training during the relevant time interval, and in column 5, the dependent variable is the total number of training days during the year. In columns 2–4 and 6–8, the three categories of training discussed in Part II are analyzed separately.

According to the specification in equation (4), the unique aspect of this analysis is the use of information on the individual's relative status in his job, that is, the ratio of his salary relative to the salaries of others performing the same job. This variable is shown as RELSTAT_{t-1} in table 5. As shown in equation (4), the effect of this variable can be measured only for individuals whose tenure in the job held at the end of $t - 1$ is at least 1 year. The effect of being in a new job in $t - 1$, that is, having tenure on this job of less than 1 year, on the likelihood of being trained during time period t is measured by the dummy variable NEWJOB_{t-1} in table 5. The other variables included in table 5 are YRSED (years of education), LOS (length of service at the firm), LOSSQ (the square of length of service), TIJ_{t-1} (tenure on the job measured at the end of $t - 1$), AGENCY (whether the source of hire was an employment agency), COLLEGE (whether the source of hire was college recruiting), WRITE (whether the source of hire was a write-in), a vector of year dummies, and a vector of broad occupational dummies.[15]

Our main interest is in the variables RELSTAT_{t-1} and NEWJOB_{t-1}. In column 1, the overall incidence equation, neither variable is significant. However, when training is separated into the three categories, some interesting findings emerge. In the cases of "core" training and technical training, the results show that individuals whose salaries at the end of time period $t - 1$ are high relative to those of their peers are more likely to receive training during the next 12 months, although the effect for core

[14] In col. 6, where the dependent variable is days of core training, OLS results are reported because the Tobit estimation procedure failed to converge.

[15] The excluded category for source of hire is advertisements.

Table 5
Determinants of Training

	Incidence				Days of Training			
	All (1)	Core (2)	Empdev (3)	Tech (4)	All (5)	Core (6)	Empdev (7)	Tech (8)
$RELSTAT_{t-1}$†	.005	.013	−.011	.008	−.003	.001	−.022	.008
	(1.55)	(1.46)	(−3.33)	(2.60)	(−.51)	(1.46)	(−3.27)	(1.08)
$NEWJOB_{t-1}$	.30	1.53	−1.28	.712	−.59	.15	−2.55	.639
	(1.09)	(1.80)	(−4.27)	(2.55)	(−.93)	(1.72)	(−4.08)	(.93)
YRSED	.022	.087	−.040	.056	.078	.005	−.072	.142
	(2.25)	(2.69)	(−3.65)	(5.50)	(3.41)	(1.53)	(−3.08)	(5.66)
LOS	−.043	.089	−.022	−.038	−.096	.003	−.054	−.084
	(−5.69)	(2.98)	(−2.44)	(−4.91)	(−5.52)	(1.17)	(−2.96)	(−4.45)
LOSSQ	.000	−.004	−.001	.001	.000	−.000	−.002	.001
	(.75)	(−3.81)	(−2.99)	(2.72)	(.096)	(−2.59)	(−3.19)	(2.40)
TIJ_{t-1}	−.03	.065	−.050	−.018	−.053	.007	−.086	−.039
	(−2.33)	(1.64)	(−3.29)	(−1.26)	(−1.71)	(1.67)	(−2.75)	(−1.16)
AGENCY	−.273	−.986	−.083	−.423	−.932	−.090	−.213	−1.29
	(−3.08)	(−3.04)	(−.92)	(−4.48)	(−4.65)	(−3.25)	(−1.13)	(−5.58)
COLLEGE	.046	−.211	−.174	.234	−.039	−.020	−.545	.533
	(.71)	(−1.13)	(−2.57)	(3.66)	(−.27)	(−1.03)	(−3.87)	(3.40)
WRITE	−.008	−.122	−.137	.172	−.072	−.019	−.378	.371
	(−.14)	(−.72)	(−2.21)	(2.91)	(−.54)	(−1.05)	(−2.92)	(2.56)
Y87	−.386	−4.69	.067	−.577	−.759	−.156	.122	−1.04
	(−7.02)	(−4.66)	(1.08)	(−10.03)	(−5.92)	(−9.01)	(.94)	(−7.40)
Y88	−.087	−.69	.373	−.339	−.240	−.072	.953	−1.05
	(−1.63)	(−4.17)	(6.37)	(−6.24)	(−1.96)	(−4.30)	(7.78)	(−7.81)
Y89	.314	−.182	.399	.205	1.26	−.027	1.02	.957
	(5.92)	(−1.33)	(7.04)	(3.99)	(10.86)	(−1.64)	(8.64)	(7.71)
Intercept	.258	−5.62	1.72	−2.14	1.59	.048	3.39	−4.20
	(.76)	(−5.55)	(4.72)	(−6.14)	(2.06)	(.45)	(4.45)	(−4.91)
Log likelihood	−15,481.1	−2,422.1	−13,647.3	−15,217.3	−23,534.6	. . .	−14,271.9	−17,381.0

NOTE.—All regressions include a vector of occupation dummies. t-Statistics are shown in parentheses. Sample size = 12. Columns 1–4 are binomial logit estimates; 5, 7, and 8 are Tobit estimates; 6 is ordinary least squares.

† Relative status in job held in $t - 1$. Equals zero if $NEWJOB_{t-1} = 1$.

training is not statistically significant. Employee development training, however, appears to be remedial; the lower an individual's RELSTAT_{t-1}, the more likely he is to receive this type of training.

The fact that RELSTAT_{t-1} is significant in the training equations and has a different effect depending on the nature of the training program rejects the hypothesis that RELSTAT_{t-1} is used by the firm simply to award merit pay increases. We find that there are two types of training in the firm, "career advancement" and "remedial," and that RELSTAT_{t-1} serves as a meaningful indicator for targeting individuals for the appropriate type of training. It was hypothesized that individuals who are new to their jobs are more likely to receive training; this is true for "core" training and technical training, but the opposite holds for employee development training. When the dependent variable is days of training rather than incidence, we find similar patterns when comparing the results for the different types of training. Individuals who score high on RELSTAT_{t-1} or who are new to their jobs participate in more days of core training, while employee development training again appears to be remedial. In the case of technical training, however, no significant results are obtained.

These findings demonstrate the importance of modeling the determinants of company training. The evidence is that this company provides two types of training: (1) training (core and technical) that is awarded to those individuals who stand out relative to their peers and (2) "remedial" training (employee development) that is targeted to those individuals who have low relative status in their jobs. In the next section, estimates of the impact of these two types of training on an employee's wage growth are presented.

B. The Effects of Training on Wages

Table 6 presents the results of estimating equation (3) using the predicted values for receipt of training and days of training from equation (4). When training is measured as a dummy variable in equation (4), all of the standard errors in equation (3) are corrected through the use of the Murphy and Topel (1985) procedure. The regular two-stage least-squares procedure is used when training is measured in days in equation (4). Although the specification in equation (3) indicates that years of education and length of service should not be included in the equation since they drop out as a result of first-differencing, the wage growth equation was estimated with and without these variables to allow for their possible impacts on wage growth. The variable CHGLOSSQ shown in table 6 is the first-differencing of the squared length of service term, that is, $\text{LOSSQ}_t - \text{LOSSQ}_{t-1}$.

The coefficients on PREDTRN in columns 1 and 5 show a positive and significant impact of overall incidence and days of training on wage growth. When training is separated into the three categories, we find that core training has a significant effect when the less restrictive specification that deletes YRSED and LOS is used (see panel II). Employee development

Table 6
The Effects of Training on Salary Growth

	Incidence				Days of Training			
	All (1)	Core (2)	Empdev (3)	Tech (4)	All (5)	Core (6)	Empdev (7)	Tech (8)
I. Instrumental variable estimates:								
PREDTRN*	.106	.019	.096	.037	.016	−.031	.016	.013
	(3.83)	(.58)	(6.48)	(3.16)	(5.16)	(−2.03)	(3.27)	(4.90)
YRSED	−.0004	.0001	.001	−.0003	−.001	.0003	.0003	−.0005
	(−1.13)	(.56)	(3.04)	(−1.19)	(−1.76)	(1.29)	(1.27)	(−1.79)
LOS	−.0015	−.0024	−.0017	−.0023	−.001	−.003	−.002	−.002
	(−5.50)	(−19.18)	(−7.17)	(−17.31)	(−3.44)	(−16.93)	(−7.90)	(−9.90)
CHGLOSSQ	.0007	.0007	.0007	.0007	.0004	.001	.001	.001
	(10.76)	(10.90)	(11.01)	(10.79)	(3.33)	(10.06)	(8.05)	(6.14)
II. Instrumental variable estimates:								
PREDTRN*	.205	.098	.150	.079	.023	.076	.046	.026
	(15.29)	(3.04)	(13.78)	(8.32)	(11.54)	(4.35)	(9.41)	(10.09)
CHGLOSSQ	.0005	−.0004	.0001	−.0002	.0000	−.0004	.0001	−.0002
	(6.98)	(−17.48)	(1.61)	(−5.98)	(.35)	(−11.86)	(1.43)	(−5.56)
III. Ordinary least squares (same specification as panel I):								
TRAIN	.01	.01	.005	.004	.001	.002	.001	.0003
	(6.02)	(3.70)	(5.37)	(4.52)	(4.07)	(3.51)	(3.82)	(1.46)

NOTE.—See the text for a complete explanation of the estimation procedure. All equations include a vector of occupation dummies and a vector of year dummies. Corrected t-values are given in parentheses. N = 12,004.
* Predicted value of incidence or days of training calculated from first-stage estimates shown in table 5.

training and technical training both have positive and significant impacts on wage growth even when YRSED and LOS are included in the wage growth equation. These results are important because they show that training in this company has a positive and significant effect on employee wage growth even after controlling for unobserved heterogeneity. Ordinary least squares estimates of the effects of training on wage growth, shown as the variable TRAIN in panel III of table 6, demonstrate the impact of heterogeneity bias. Since core training is provided to "stars" in the company, the ordinary least squares estimates show a positive and significant effect of core training on wage growth that disappears when the two-step estimation method is used. Alternatively, employee development training, which is remedial in nature, has a much larger impact on wage growth when the two-step procedure is used.

In order to test the alternative hypothesis that employees with high $RELSTAT_{t-1}$ have higher wage growth irrespective of the training that they receive, equation (3) was reestimated including $RELSTAT_{t-1}$ in the equation. The results are shown in table 7. While $RELSTAT_{t-1}$ has a positive and significant effect on wage growth, the significant effects of training on wage growth that were observed in table 6 (with the exception of core training in panel II) remain. Taken together, the results in tables 5, 6, and 7 show that $RELSTAT_{t-1}$ is used by the firm to select employees for career advancement training or remedial training, that training does indeed lead to higher wage growth, and that $RELSTAT_{t-1}$ also has a direct (i.e., not operating through training) positive effect on wage growth.

The results presented in table 7 can be used in conjunction with information on the costs of training at the company to calculate the rates of return the company earns on the training it provides. From company documents I was able to calculate the per-participant direct costs of a day of training, which includes the salaries of the trainers and the costs of materials, room, and board. The indirect costs of training were calculated from data on the salaries of the trainees. Direct and indirect costs were then summed to determine the per-participant training cost. On average, during the 1986–90 time period, it cost the company $1,440 to provide 1 day of training to an employee.

The company's returns from training are the productivity gains that training produces. The magnitude of these productivity gains can be inferred from the wage gain estimates shown in table 7. A conservative estimate is that the productivity gains equal the wage gains. This assumption is consistent with a company having a value added/wage ratio equal to one.[16]

[16] Since value added is defined as the value of sales minus labor costs, a value added/labor cost ratio of 1 implies a sales/labor cost ratio of 2. A value added/labor cost ratio of 1 means that a $1 increase in labor costs raises value added by $1 and sales by $2. Most companies have value added/labor cost ratios that exceed 1.

Table 7
The Effects of Training on Salary Growth, Controlling for RELSTAT$_{t-1}$

	Incidence				Days of Training			
	All (1)	Core (2)	Empdev (3)	Tech (4)	All (5)	Core (6)	Empdev (7)	Tech (8)
I. Instrumental variable estimates:								
PREDTRN*	.06	−.017	.066	.047	.018	−.069	.016	.016
	(2.61)	(−.53)	(4.71)	(3.95)	(5.44)	(−3.44)	(3.35)	(5.58)
RELSTAT$_{t-1}$	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
	(10.89)	(13.58)	(9.00)	(12.79)	(9.57)	(10.12)	(12.03)	(11.27)
YRSED	.0001	.0004	.0009	−.0003	−.0005	.0007	.0005	−.0005
	(.18)	(1.80)	(3.26)	(−.94)	(−1.31)	(2.27)	(2.10)	(−1.48)
LOS	−.0018	−.0023	−.0017	−.0021	−.0007	−.0025	−.0016	−.0016
	(−7.51)	(−18.02)	(−8.86)	(−15.11)	(−2.25)	(−13.13)	(−7.07)	(−7.76)
CHGLOSSQ	.0007	.0001	.0007	.0007	.0003	.0007	.0006	.0004
	(10.23)	(10.23)	(10.34)	(10.54)	(2.34)	(7.98)	(7.29)	(4.69)
II. Instrumental variable estimates:								
PREDTRN*	.183	.064	.125	.087	.023	.044	.043	.027
	(10.76)	(.96)	(10.44)	(6.57)	(11.48)	(3.10)	(9.23)	(10.23)
RELSTAT$_{t-1}$	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
	(5.84)	(14.58)	(7.63)	(11.92)	(8.65)	(12.33)	(7.83)	(9.75)
CHGLOSSQ	.0004	−.0004	.00003	−.0002	.0000	−.0004	.0001	−.0002
	(4.64)	(−16.63)	(.41)	(−4.05)	(−.55)	(−14.27)	(1.31)	(−4.80)

NOTE.—See the text for a complete explanation of the estimation procedure. All equations include a vector of occupation dummies and a vector of year dummies. Corrected t-values are given in parentheses. N = 12,004.
* Predicted value of incidence or days of training calculated from first-stage estimates shown in table 5.

Previous studies have found that productivity increases from training and reduced turnover are at least double the wage increases that employees receive.[17] The latter findings are consistent with a value added/wage ratio that exceeds 1, which is in fact the case for the company under study. Using the conservative, lower-bound estimate that productivity gains equal wage gains results, therefore, in a lower-bound estimate of the company's rate of return on employee training.

I calculated rates of return for each category of training using the coefficients on PREDTRN from columns 6, 7, and 8 in panel I of table 7. Three alternative assumptions were used in making these calculations: (1) skills do not depreciate; (2) skills depreciate at the rate of 10% per year, implying a half-life of 5.4 years for the investment; and (3) skills depreciate at the rate of 20% per year, implying a half-life of 2.9 years. The results are shown in table 8. Assuming that skills depreciate at the rate of 10% per year, the company earns rather high rates of return, for example, 34.6% on employee development training and 36.6% on technical training. Similar calculations for national data sets that are not restricted to one company and where individuals self-report training events have found rates of return of 16.4% (PSID data set) and 18.6% (NLSY data set), when depreciation rates of 10% are assumed (Mincer 1991). In this company, rates of return of this magnitude would require an assumption of a much higher depreciation rate, as shown by the calculations based on a 20% depreciation rate. There are several possible explanations for the disparity in rates of return between this company and the national data sets. One possibility is that training in this company does indeed have a larger payoff than the amount typically calculated for a national sample of employees. The second reason may be that, unlike the company database, national data sets typically do not include information on very short training programs.[18] It is possible that these short training programs have higher rates of return than the longer training spells that the national data sets measure.

C. Effects of Training on Job Performance

Finally, this data set provides the unique opportunity to measure the impact of training on job performance. At the end of 1988, the company began to use a system of relative performance measures whereby individuals are ranked relative to their peers. According to the company's policies, managers are expected to differentiate clearly among employees in order to recognize those whose performance is far superior to that of their peers. For groups of at least 100 employees in the company, managers are expected to follow a targeted distribution of ratings. There are seven levels used in

[17] See Blakemore and Hoffman (1988) and Barron, Black, and Loewenstein (1989).

[18] See n. 3 above for a description of this problem as it pertains to the NLSY database.

Table 8
Rates of Return on 1 Day of Training in Each Category (%)

	All (1)	Core (2)	Empdev (3)	Tech (4)
Assuming depreciation rate = 0%	57.6	Negative	49.5	51.8
Assuming depreciation rate = 10%	41.8	Negative	34.6	36.6
Assuming depreciation rate = 20%	26.1	Negative	19.6	21.4

NOTE.—See the text for the method of calculation.

the performance evaluation system. Performance evaluations are done in January of each year for individuals who have been in their jobs for at least 1 year. If the job the individual held in January 1989 began in January 1988 or earlier, his performance score for that job is recorded on the 1989 personnel record. Similarly, if the job held in January 1990 began in January 1989 or earlier, the performance score is recorded on the 1990 record. Therefore, if an individual did not change jobs between January 1988 and January 1990, the change in his job performance can be calculated by using the performance scores that are recorded in January 1989 and January 1990. Some of these individuals may have participated in training programs during calendar year 1989. It is therefore possible to see if individuals who received training during 1989 also improved their performance ratings on the same job.

The impact of training on performance ratings can serve as an estimate of the impact of training on true performance. The estimate is likely to be biased downward, however, because the variation in performance scores will be less than the variation in true performance. An additional problem relates to the fact that this analysis must be restricted to individuals who did not change jobs between January 1988 and January 1990, thereby eliminating the possible impact of training on upward mobility within the firm. Restricting the analysis to nonjob-changers underestimates the true impact of training on "performance" in the firm. It is impossible, however, to include job changers in the analysis because a performance rating in one job cannot be compared to a performance rating in a different job.

First, the wage growth equation, equation (3), is estimated on the sample of employees who did not change jobs between 1988 and 1990.[19] Although the estimated coefficient on PREDTRN was .106 in the full sample (see table 6), it is only .029 with a *t*-value of .74 in this sample. The weak effects of training on wage growth for individuals who did not change jobs over a 2-year period occurs because a large part of training's positive impact on within-firm wage growth is due to upward mobility within the firm.

[19] The two-step procedure discussed earlier is also used here. The training incidence equation is first estimated, and predicted values of training are used in the wage growth equation.

Equation (5), the change in performance scores between January 1989 and January 1990, is estimated using multinomial logit. The change in performance scores is categorized into one of the following three choices: (1) the performance score increased between 1989 and 1990, (2) the performance score remained unchanged between 1989 and 1990, and (3) the performance score decreased between 1989 and 1990. Predicted values of the probability of receiving training and the days of training received during 1989 are calculated from first-stage estimates of equation (4) on the sample of individuals who did not change jobs between January 1988 and January 1990, and these predicted values are used in the second-stage multinomial logit equations.

Table 9 shows the coefficients on PREDTRN from the multinomial logit equations; the other variables included in these equations are YRSED, LOS, and a vector of occupation dummies. The excluded performance change category is the probability of a decrease in the performance score. The results appear to be very sensitive to the aggregation or separation of the different types of training, and the strongest results are in column 1, where the impact of receiving any type of training is estimated. The results in that column show that individuals who received training during 1989 were significantly more likely to receive increases in their performance ratings between January 1989 and January 1990. The receipt of training also increased the probability that an individual's performance score did not decline. These findings of strong positive effects of training on job performance are rather remarkable considering the two downward biases discussed earlier, that is, that the analysis is restricted to individuals who remained in the same job between January 1988 and January 1990, and that the performance scores are an imperfect measure of true job performance. In spite of these factors, we do find that training did lead to an increase in the measured performance of this sample of employees, strengthening our confidence in the earlier results where wages were used as the proxy for the employee's productivity.

V. Conclusions

In recent years, the availability of data on training, largely collected in national surveys of individuals, has allowed researchers to directly analyze the link between on-the-job training and wages. Unfortunately, this line of research has been plagued by a number of measurement problems, among them the difficulty that individuals have in recalling the type and amount of training they received as well as the comparability of training events across individuals working in different organizations. In this article, a unique data set collected from the personnel records of a large company is used to study the relationship between on-the-job training and worker productivity. By using a company database, I was able to avoid the measurement problems that typify national data sets. First, since all training

Table 9
Multinomial Logit Estimates of the Impact of Training on the Probability of a Change in Job Performance for Employees Who Did Not Change Jobs between 1988 and 1990 (*t*-Values Are in Parentheses)

	Incidence				Days of Training			
	All (1)	Core (2)	Empdev (3)	Tech (4)	All (5)	Core (6)	Empdev (7)	Tech (8)
Performance score remains the same:								
PREDTRN*	.10	N.A.†	.01	−.034	.035	.217	.175	.022
	(1.41)		(.11)	(−.52)	(1.18)	(.71)	(1.98)	(.55)
Performance score increases:								
PREDTRN*	.185	N.A.†	−.049	−.038	.012	−.286	.087	.01
	(2.53)		(−.48)	(−.35)	(.40)	(−.93)	(.95)	(.22)

NOTE.—All regressions include YRSED, LOS, and a vector of occupation dummies and are estimated via the method of maximum likelihood multinomial logit. The coefficients reported are the change in the probability due to a 1-unit change in PREDTRN. N = 1,478. N.A. = not applicable.
* Predicted value of incidence or days of training from first-stage estimates of eq. (4).
† Estimates did not converge.

spells and their exact durations are recorded on the individual's personnel record, precise measures of training duration, even for programs that last only a few days, are obtained. Second, there is no bias resulting from definitions of training varying across diverse firms because all of the individuals are trained by the same firm.

The analysis of the company database produced several important findings that will influence the way economists think about the relationships between on-the-job training and wages and job performance. First, there is evidence that formal company training spells tend to be rather brief, lasting only a few days. National data sets that report training duration in terms of weeks, not days, fail to capture this important component of training. Second, the analysis showed how information contained in a company database could be used to correct for the impact of heterogeneity bias in the estimation of training's effect on wage growth. Evidence was presented that, in this company, individuals were selected for job training based on their relative status in a job, that is, how their salaries compared to other individuals with identical education, years of company service, and time on the current job. The implication of this finding is that research on the relationship between investments in training and wage growth must consider why some employees receive training and others do not. Third, it was shown that, even after the selection bias in assignment to training was eliminated, individuals in this company who received training experienced significant increases in wage growth. Fourth, the company database provided the unique opportunity of using data on performance ratings as another proxy for productivity (apart from wages). Training was found to have a positive and significant effect on job performance, thereby confirming the robustness of the relationship between training and productivity.

This article has shown that an analysis of training, wages, and performance that eliminates the influence of company-specific experiences on employees by studying one company confirms the findings of previous research that relied on publicly available data sets. The main finding in this article is that formal training does indeed increase wage growth and job performance, even when selection bias in assignment to training programs is eliminated. While the results may not generalize to other firms, this study has shown that economists can learn important lessons from analyzing a company database. Future research in this area should focus on collecting more comprehensive data from companies that will include information on both formal and informal training, as well as more precise measures of worker productivity.

References

Barron, John M.; Berger, Mark C.; and Black, Dan A. "New Evidence on the Measurement of Training: Matched Employee and Employer Responses." Mimeographed. Lexington: University of Kentucky, January 1994.

Barron, John M.; Black, Dan A.; and Loewenstein, Mark A. "Job Matching and On-the-Job Training." *Journal of Labor Economics* 7 (January 1989): 1–19.

Bartel, Ann P., and Sicherman, Nachum. "Technological Change and On-the-Job Training of Young Workers." Mimeographed. New York: Columbia University Graduate School of Business, March 1994.

Blakemore, Arthur, and Hoffman, Dennis. "Seniority Rules and Productivity: An Empirical Test." Mimeographed. Tempe: Arizona State University, September 1988.

Brown, James N. "Why Do Wages Increase with Tenure?" *American Economic Review* 79 (December 1989): 971–91.

Holzer, Harry. "The Determinants of Employee Productivity and Earnings." *Industrial Relations* 29 (Fall 1990): 403–22.

Jovanovic, Boyan. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (October 1979): 972–90.

Lazear, Edward. "Agency, Earnings Profiles, Productivity and Hours Restrictions." *American Economic Review* 71 (September 1981): 606–20.

Lillard, Lee, and Tan, Hong. "Private Sector Training: Who Gets It and What Are Its Effects?" Rand Monograph R-3331-DOL/RC. Santa Monica, CA: RAND, March 1986.

Loewenstein, George, and Sicherman, Nachum. "Do Workers Prefer Increasing Wage Profiles?" *Journal of Labor Economics* 9 (January 1991): 67–84.

Lynch, Lisa. "Private Sector Training and the Earnings of Young Workers." *American Economic Review* 82 (March 1992): 299–312.

Medoff, James, and Abraham, Katherine. "Are Those Paid More Really More Productive? The Case of Experience." *Journal of Human Resources* 16 (Spring 1981): 186–216.

Mincer, Jacob. *Schooling, Experience and Earnings.* New York: Columbia University Press, 1974.

———. "Job Training: Costs, Returns, and Wage Profiles." In *Market Failure in Training? New Economic Analysis and Evidence on Training of Adult Employees,* edited by David Stern and Jozef M. M. Ritzen. Berlin: Springer-Verlag, 1991.

Murphy, Kevin M., and Topel, Robert H. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3 (October 1985): 370–79.

Salop, Joanne, and Salop, Steven. "Self-Selection and Turnover in the Labor Market." *Quarterly Journal of Economics* 90 (November 1976): 619–27.

Sicherman, Nachum. "The Measurement of On-the-Job Training." *Journal of Economic and Social Measurement* 16 (January 1991): 1–10.

Training Magazine. Minneapolis: Lakewood Publications, October 1990.

U.S. Congress, Office of Technology Assessment. *Worker Training: Competing in the New International Economy.* OTA-ITE-457. Washington, DC: U.S. Government Printing Office, 1990.