# Measuring the Employer's Return on Investments in Training: Evidence from the Literature

ANN P. BARTEL*

Three components of the literature on measuring the employer's rate of return to investments in employee training are reviewed: (1) studies that use large samples of firm-level or establishment-level data collected through mail or phone surveys, (2) studies that use data from one or two companies to conduct an "econometric" case study, and (3) company-sponsored case studies. The strengths and weaknesses of each of these approaches are evaluated and the estimated *returns on investments* (ROIs) are compared. The analysis indicates that the employer's return on investments in training may be much higher than previously believed. In order to obtain accurate information on the employer's ROI from training, researchers should be encouraged to gain access to company databases and to supplement them with data-gathering efforts to collect information on variables needed to isolate the effect of training.

## Introduction

Although American businesses budgeted $58.6 billion for formal training in 1997 (Lakewood Publications, 1997), there is no general consensus on the rate of return that employers earn on this investment. In the literature, there are two approaches to measuring this return. The first approach uses data on a large sample of firms and compares the experiences of firms that invest in training with those which do not. In the second approach, typically labeled the *case study method*, detailed data from one company are gathered in order to estimate the costs and returns from the

*Columbia University Graduate School of Business and NBER. E-mail: *apb2@columbia.edu*. Funding for this research was provided by the U.S. Department of Labor under Contract No. F-4956-5-00-80-30. I gratefully acknowledge helpful comments provided by several anonymous referees.

INDUSTRIAL RELATIONS, Vol. 39, No. 3 (July 2000). 
Published by Blackwell Publishers, 350 Main Street, Malden, MA 02148, USA, and 108 Cowley Road, Oxford, OX4 1JF, UK.

company's training program. The advantage of the second approach is that it has the potential to accurately control for all other factors besides training (e.g., worker and firm characteristics) that may influence productivity. The disadvantage is that the results from a case study may not generalize to other companies. While the first approach avoids the generalization problem, it can suffer from lack of insufficient data to accurately measure a rate of return on investment.

It is clearly important for companies to have accurate measures of the rate of return on investments (ROI) in employee training, for this is what guides their human capital investment decisions. If the expected ROI is underestimated, employers will underinvest, whereas if it is overestimated, employers will overinvest. Similarly, knowledge of the rate of return on company investments in training is important for government policymakers who may be interested in allocating government resources to subsidize private investment.

This article reviews the literature on measuring the employer's return on investing in formal company training. This is in contrast to the rate of return that the employee may calculate or the rate of return that may accrue to society as a whole.[1] Calculating the employer's return is complicated by the fact that employers and employees may share in the costs and returns of training. Recent work by Loewenstein and Spletzer (1998) that uses the 1988–1991 National Longitudinal Surveys of Youth documents, however, that employers pay for nearly all spells of formal company training and a large portion of what appears to be general training.[2] Hence, by focusing on formal company training, it can be assumed that the costs of the training are borne almost entirely by employers. Loewenstein and Spletzer's finding regarding the employer's cost share explains why other researchers have found that the effect of an hour of company training on productivity growth is about five times as large as its effect on wage growth (Barron et al., 1989, 1993; Bishop, 1991), i.e., that employers reap almost all the returns to company training.

In this article I review and contrast the findings from case studies of single companies and articles that use econometric techniques to analyze data from large numbers of firms. Most of the studies reviewed here use a direct measure of productivity growth rather than wage gains to measure

[1] Given the inherent difficulties in measuring informal training, this article focuses on the rate of return to formal training only.

[2] It is possible that an employer who incurs the explicit cost of training may pass it on to the worker in the form of a lower wage. Loewenstein and Spletzer (1998) show, however, that wages are not reduced during the training period.

the benefit from training.[3] The purpose of this review is twofold: (1) to clarify our understanding of the employer's ROI in company training (What can we learn from the current literature and how reliable are the numbers for making policy decisions regarding private and public resource allocation?) and (2) to suggest directions for future research that will improve the accuracy and reliability of the measured rates of return. Previous reviews of the training literature have analyzed other issues[4]; there also have been a number of studies of the relationship between other human resource management practices and firm performance, although many of these have not had sufficient data to calculate rates of return.[5]

The second part of this article reviews the studies that have estimated the impact of training on organizational productivity by using firm-level or establishment-level data collected through mail or phone surveys. It is shown that these studies do not provide much guidance on the question of the employer's rate of ROI in training because they typically lack data on training costs, they may not accurately model a firm's unique production process, and they may not sufficiently account for the endogeneity of the training decision. The third part reviews the articles that have addressed the first two problems by applying an econometric framework to data from one or two companies and conducting an "econometric case study." While these studies are an improvement over the earlier approach, their results may not be entirely free of endogeneity bias. The fourth part considers what we can learn from a third source of information on the employer's rate of return on training investments, i.e., the companies' own evaluations of their training programs. This section shows that a review of the human resource management literature for the time period 1987 through 1997 uncovered a total of only 16 company case studies in which the rate of return on training investments was measured. The vast majority of these case studies are, unfortunately, plagued by a number of serious methodologic flaws, such as inappropriate evaluation design, lack of attention to selection bias, insufficient controls for other variables that influence performance, monitoring the impact of training for a relatively

---

[3] While it is clearly preferable to use productivity gains rather than wage gains, the estimated ROI to the employer may still be imperfect if some of the benefit from training is captured by employees over time.

[4] See Goldstein (1992) for a review of the literature on the design of training systems and appropriate evaluation criteria and Bishop (1997) for a review of the literature on the incidence, determinants, and impacts of training.

[5] A number of journals have devoted special issues to the topic of measuring the relationship between human resource management practices and firm performance. See the July 1996 issue of *Industrial Relations*, the August 1996 issue of *Academy of Management Journal*, and the Fall 1997 issue of *Human Resource Management*. None of the articles in those issues conducted an analysis of the employer's ROI in training.

short time, and using self-reports from the trainees about the productivity gains from training. Two case studies that are well conceived are reviewed, and their results are compared with the findings from the econometric case studies. The main conclusion of this article, provided in the final section, is that a company's ROI in employee training may be higher than previously thought, and companies should be encouraged to use internal databases to calculate the ROI on their training investments.

## Econometric Analyses of Large Samples of Firms

A number of studies have estimated the impact of training on organizational productivity by using firm-level or establishment-level data collected through mail or phone surveys. The studies of this type that are most frequently cited are Bishop (1991), Bartel (1994), Holzer et al. (1993), Black and Lynch (1996a, 1996b), and Tan and Batra (1995). Huselid's (1995, 1996) analysis of high-performance work practices also takes the same approach of collecting survey data from a large sample of heterogeneous firms. As I describe below, these studies are unable to calculate the rate of return on investing in training because of the absence of reliable data on costs of training. In addition, the extent to which the reported results can truly be interpreted as productivity impacts depends on the authors' success in correcting for the endogeneity of the training decision; in some cases, positive productivity impacts disappear after the endogeneity correction. The main attributes and findings of these studies are summarized in Table 1.

Bishop (1991) used data on 2594 employers from the Employment Opportunity Pilot Projects (EOPP) Employer Survey sponsored by the National Center for Research in Vocational Education. In phone interviews conducted in the spring of 1982, the employers were asked to select "the last new employee your company hired prior to August 1981 regardless of whether that person is still employed by your company." For that new hire, the employer estimated how much time was spent in the first 3 months on formal training. The employer also reported on the productivity of the typical individual hired into the job after 2 weeks, during the next 11 weeks, and at the end of 2 years with the firm. The rating was made on a scale of 0 to 100, where 100 equaled the maximum productivity that any employee in that position could obtain. Using these data, Bishop estimated the impact of training in the first 3 months on the job on the 2-year growth in the typical worker's productivity. He estimated costs of training using the trainee's time costs and an arbitrary adjustment factor for the other costs of training. The estimated marginal rate of return for 100 hours of training ranged from 11 percent for the linear

TABLE 1
ECONOMETRIC ANALYSES OF LARGE SAMPLES OF FIRMS

| Author | Dataset | Response rate | Sample size | Performance measure | Findings |
|---|---|---|---|---|---|
| 1. Bishop | EOPP (1982) | 75% | 2594 businesses | Productivity rating | ROI on 100 hours of new hire training ranged from 11% to 38% |
| 2. Bartel | Columbia HR Survey (1986) | 6.5% | 155 manufacturing businesses | Value-added per worker | Implementation of formal training raised productivity by 6% per year |
| 3. Holzer et al. | Survey sent to Michigan firms applying for state training grants | 32% | 157 firms | Scrap rate | Doubling of worker training reduced scrap rate by 7%, using fixed-effects model; this is worth $15,000 |
| 4. Black and Lynch | EQW National Employers Survey (1994) | 72% | 617 manuf. establishments | Net sales | Percentage of formal training that occurs off the job has significant effect in cross section but no effect on the establishment-specific residual |
| 5. Tan and Batra | World Bank Survey | Random sample | 300–56,000 per country | Value-added | Predicted training has positive effect on value-added; effects range from 2.8% to 71% per year |
| 6. Huselid | 1992 survey of human resource practices | 28% | 968 firms | Tobin's q and gross rate of return on capital | High-performance practices had significant effect in cross section that disappeared in fixed-effects model |

specification of the productivity equation to 38 percent for the logarithmic specification. The reliability of these estimates depends on the accuracy of Bishop's assumptions regarding the cost of training and the correlation between training intensity during the first 3 months and training hours during the rest of the 2-year period, as well as the accuracy of the subjective estimates of productivity.

In Bartel (1994), I used data from a 1986 Columbia Business School survey of 495 business lines to estimate the impact of formal training programs on productivity.[6] I first estimated a 1983 labor productivity

[6] For those firms which operate only one business line, a *business line* corresponds to the company. When a parent company operates several business lines, a *business line* generally corresponds to a division of a company. Although the response rate for the Columbia survey was only 6.5 percent, the sample closely matched the industrial distribution of all 1986 Compustat II business lines. The analysis was restricted to businesses in the manufacturing sector because of the availability of data on the costs of purchased materials. This resulted in a sample of 155 businesses.

equation and calculated the residual for each business, i.e., the extent to which the business's output deviated from the output that would be expected given its input use and its industry. In the next step, restricting the analysis to those businesses which did not have training programs as of 1983, I estimated an equation in which the probability of implementing a training program after 1983 was regressed on the residual from the 1983 labor productivity equation and the age of the business. The main finding was that businesses that had negative residuals in 1983 were more likely to introduce training after 1983; in other words, if a business had an output level in 1983 that was below what would be expected given its input use and its industry, it was more likely to introduce a formal training program. Businesses that implemented formal training programs after 1983 experienced an 18 percent increase in productivity between 1983 and 1986 (i.e., a 6 percent annual increase) compared with businesses that did not. Implementation of other human resource policies (e.g., job design, performance appraisal, employee involvement) during the same time period did not have a productivity-enhancing effect, rejecting a Hawthorne effect explanation for the findings. Data on the cost of the training programs were largely missing, thereby precluding a cost-benefit analysis.

Holzer et al. (1993) used data on firms that applied for training grants under the Michigan Job Opportunities Bank Upgrade in 1988 and 1989.[7] Firms that responded to Holzer's 1990 survey provided retrospective data for 1987, 1988, and 1989 on annual hours of training per employee, the scrap rate (a measure of output quality), the value of sales, employment, and other human resource policies.[8] Using a model that controlled for the effects of fixed unobservable firm characteristics, Holzer et al. estimated the impact of hours of training on the scrap rate. They found that a doubling of worker training in any year produced a contemporaneous reduction in the scrap rate of 7 percent that was worth about $15,000 per year, but half of this effect disappeared in the next year, and presumably even more would be gone in subsequent years. Unfortunately, Holzer et al. did not collect information from the firms about their training expenditures, so they were unable to conduct a cost-benefit analysis. Another limitation of this article is that the data are retrospective, and it is not clear that the respondents always used objective measures from company databases to respond to the questionnaire.

[7] Eligible firms were those in the manufacturing sector with less than 500 employees and in the process of implementing some type of new technology.

[8] The survey response rate was 32 percent.

Black and Lynch (1996a, 1996b) used data collected from the National Center on the Educational Quality of the Workforce (EQW) National Employers Survey, which was administered by the U.S. Bureau of the Census as a telephone survey in August and September 1994 to a nationally representative sample of more than 3000 private establishments with more than 20 employees.[9] Black and Lynch (1996a) estimated a production function in which the dependent variable is the dollar value of sales, receipts, or shipments in calendar year 1993. They found that the number of workers trained in 1990 and 1993 had no impact on productivity, but the percentage of formal training off the job was positive and significant for the manufacturing sector, and computer training was positive and significant in the nonmanufacturing sector. Of course, the cross-sectional approach taken by Black and Lynch (1996a) does not address the problem of endogeneity, i.e., that the firm's performance level influences its decision to invest in employee training. In a follow-up paper, Black and Lynch (1996b) address this problem by restricting the analysis to the manufacturing sector and matching the establishments to the Census Bureau's Longitudinal Research Database. In this article, Black and Lynch estimate a first-difference production function (including labor, capital, and materials as regressors) for the time period 1988–1993 and use the coefficients from this equation to estimate an establishment-specific residual that is then regressed on variables measuring the establishment's human resource management practices obtained from the telephone survey. None of the training variables were significant in the second stage. Once the endogeneity issue is properly addressed, the positive relationships between training and productivity observed in the cross-sectional analysis disappear. Further, even if these effects did not disappear, Black and Lynch provided no data on training costs, so it is impossible to compare benefits against costs.

Under the auspices of the World Bank, Tan and Batra (1995) assembled a unique set of firm-level data to study training in five developing countries—Indonesia, Colombia, Malaysia, Mexico, and Taiwan.[10] For each country, Tan and Batra estimated a production function in which the dependent variable is the logarithm of value added and the independent

[9] The response rate was 72 percent, which is extremely high.

[10] Colombia, Indonesia, and Malaysia fielded surveys of manufacturing firms using a survey instrument designed by the World Bank project team. Mexico used a survey instrument developed jointly by the Secretariat of Labor and Social Welfare and the ILO, with input from the World Bank team to ensure comparability with the other country surveys. The surveys were fielded between 1993 and 1995 and oversampled large firms. Sample sizes are Colombia (500), Indonesia (300), Malaysia (2200), and Mexico (5072). The Taiwanese data are drawn from the 1986 Manufacturing Census, which contains 56,057 firms.

variables are the logarithms of labor and capital, the rate of capacity utilization, mean education, whether the firm is an exporter, whether it conducts research and development (R&D), whether it possesses foreign technology licenses, a set of two-digit industry dummy variables, and a dummy variable indicating if it provides any formal training.[11] In order to deal with the possible endogeneity of the training decision, Tan and Batra first estimated a training equation and used the predicted value for training in their productivity equation. They found that training had a positive and significant effect in all five countries. The coefficients were lowest for Taiwan (0.028) and highest for Indonesia (0.711); falling in between were the training effects for Colombia (0.266), Malaysia (0.282), and Mexico (0.444). While Tan and Batra should be commended for addressing the fact that the firm's decision to train may be determined by its productivity level, it is not clear that their system of equations has been identified properly. A number of variables that are used in the training equation and eliminated from the productivity equation arguably could belong in the productivity equation as well[12]; the result of this misspecification would be an overestimate of the true effect of training on productivity.

Huselid (1995) studied the impact of high-performance work practices by using survey data on 968 firms.[13] The survey was completed in 1992 by the senior human resources executive in the firm and solicited information about the firm's 1991 human resource policies, including the average number of hours of training received by a typical employee over the last 12 months.[14] Financial data on the firm for the time period July 1, 1991 through June 30, 1992 were obtained from Compact Disclosure, and stock price data also were gathered. These data were used to calculate two measures of financial performance: Tobin's $q$ (the market value of the firm divided by the replacement cost of its assets) and the gross rate of return on capital (cash flow divided by gross capital stock). Arguing that it is inappropriate to study the impacts of individual human resource practices because firms have systems of human resources management practices that are defined by the internal consistency of those practices, Huselid grouped the firm's human resource policies into two categories:

[11] The education variable was not available in the Taiwanese data.

[12] Examples are exports, age of the firm, multiplant status, proportion of female workers, unionization, and percentage value of automatic machinery.

[13] This sample resulted from a response rate of 28 percent. Huselid tested for selectivity bias in the responses to the survey and found that the corrected empirical results were virtually identical to the uncorrected ones.

[14] Measuring a firm's human resource management policies may be problematic if the firm operates multiple business lines that do not have identical human resource policies.

employee skills and organizational structure (i.e., selection, training, information sharing, quality of work life) and employee motivation (pay-for-performance, promotions based on merit or seniority).[15] He found that both sets of policies had significant impacts on Tobin's *q*, whereas only employee skills and organizational structure were significant for the gross rate of return on capital. Each one standard deviation increase in high-performance work practices increased cash flow by $3814 and raised the firm's market value by $18,641.

The advantage of Huselid's approach is that he focuses on profits, so the costs of implementation of high-performance work practices have been netted out. The problem with his approach is that it is cross-sectional; i.e., 1991 policies are regressed on 1991–1992 profits. A follow-up article (Huselid and Becker, 1996) reports the results of estimating the equations on a panel data set. Of the original 968 firms, 218 completed a follow-up survey in 1994 on which they reported information on 1993 policies. With these data, Huselid and Becker were able to estimate a model that accounts for the impact of fixed, unobserved firm characteristics. In this model, the results were insignificant and were only 25 to 30 percent as large as the original estimates.[16] The dramatic fall in the coefficients from the single-time-period model to the panel-data model demonstrates the importance of correcting for endogeneity.

To summarize, the econometric studies of large-sample databases do not provide much guidance on the question of the employer's rate of ROI in training. First, data on costs of training typically are not available in these datasets, thereby limiting the findings to estimates of productivity impacts. Second, by using data on heterogeneous firms, these studies may not be accurately modeling diverse production processes. Third, while all the studies (with the exception of that of Bishop) attempted to correct for the endogeneity of the training decision, we cannot be totally confident that the results are still not at least partially plagued by this bias.

## Econometric Case Studies

In order to address the problems created by studying a sample of heterogeneous firms, a number of researchers have tried to measure the employer's returns to training by applying an econometric framework to

[15] Theoretical work on incentive contracts suggests that bundles of human resource policies exist because of technical complementarities; i.e., the benefit from one policy increases if other, related policies are also used. See Milgrom and Roberts (1995).

[16] The authors attribute this to the increased importance of measurement error when panel data are used, and they argue that the panel data results should be inflated.

data from one or two companies and conducting what one might call an "econometric case study." The two examples of this approach are the work by Bartel (1995) and Krueger and Rouse (1998). In the literature on measuring the impact of high-performance work practices, the article by Ichniowski et al. (1997) uses this approach in its analysis of 36 steel finishing lines. All these studies share the common characteristic of direct collection of data from the company's personnel files and face-to-face interviews with managers to understand the wage determination and production processes and the role of training and other human resources policies at the firm. In addition, access to cost data enables these researchers to calculate the ROI in training (or a package of human resources policies). Table 2 summarizes the attributes and main findings of these studies.

In Bartel (1995), I used data from the 1986–1990 personnel records of a large manufacturing company to estimate the company's rate of ROI in training its professional employees in management skills, communications skills, and technical areas. In the sample of 19,000 employees, at least half received some formal training during each of the 5 years, and their mean days of training per year ranged from 3.3 to 4.4 days. In order to deal with the fact that the receipt of training may have been related to characteristics of the individuals, I first estimated a training incidence equation in which the likelihood of receiving training was a function of the employee's relative status in his or her job and other characteristics

TABLE 2
ECONOMETRIC CASE STUDIES

| Author/company | Employee group | Type of training | Performance measure | ROI | Comments |
|---|---|---|---|---|---|
| 1. Bartel/Large Manuf. Co. | Professional employees | Mgmt., commun., and technical skills | Wage growth and performance ratings | 49.7%[a] | Controls for selection bias and uses fixed-effects model |
| 2. Krueger & Rouse Manuf. Co. | Lower-skilled | Reading, writing, and math | Wage growth and performance awards | 7%[a] | Uses fixed-effects model |
| 3. Krueger & Rouse Service Co. | Lower-skilled | Reading, writing, and math | Wage growth and performance awards | Zero | Uses fixed-effects model |
| 4. Ichniowski et al. Steel finishing lines | Production workers | High-performance work practices | Uptime of line | N.A. | Switching to high-performance work systems raised monthly profits by $27,900, using fixed-effects model |

[a]Based on assumption that skills depreciate at the rate of 5% per year.

such as education, years of service on the job and at the company, and source of hire. Relative status was reported in the company database as the employee's salary divided by the average salary of other employees in the same job. I argued that the individual's relative status in his or her 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 $t$. If training is not remedial, 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, an individual with lower relative status would be selected for training in order to bring his or her skills up to the "average" level. I found strong evidence that assignment to training was indeed based on the individual's relative status. Calculating the predicted values of training from the training incidence equation and using these as regressors in first-difference wage equations that eliminated person-specific fixed effects, I found that each day of training raised wages by 1.8 percent. In addition, individuals who received training experienced an increase in their job performance scores, thereby confirming the robustness of the relationship between training and productivity.[17]

Since I had access to company records, I was able to calculate the cost of a day of training, combining information on both direct costs and the cost of time taken away from work. By assuming that the company had a value-added/wage ratio equal to one, I estimated that the company's productivity gains from training were of the same magnitude as the estimated wage gains.[18] Rates of return were calculated based on various assumptions regarding the rate at which skills depreciate per year. For example, with an annual depreciation rate of 10 percent, the internal rate of return was 41.8 percent. Doubling the depreciation rate to 20 percent produced an internal rate of return of 26.1 percent, whereas halving it to 5 percent resulted in an internal rate of return of 49.7 percent.

In a similar study, Krueger and Rouse (1998) collected data from two New Jersey companies, one in manufacturing and one in the service sector, that sought help from a community college in designing on-site training for lower-skilled workers. The training covered basic education in reading, writing, and math. Krueger and Rouse collected pretraining and posttraining wage data on the workers as well as detailed personal

[17] This part of the analysis was restricted to individuals who did not change jobs over a 2-year period in order to be able to compare their performance ratings on the same job.

[18] This is actually a very conservative assumption, since most companies have value-added/labor costs ratios that exceed one. Also, as discussed in the Introduction, previous research has found that the effect of an hour of training on productivity growth is five times as large as its effect on wage growth.

characteristics and information on turnover, absenteeism, and performance awards. As in my study, they estimated the impact of training on wages by using a model that eliminated the impact of any unobserved individual characteristics. In the manufacturing company, they found that the return to training was about 0.5 percent, whereas there was no effect of training in the service sector. In order to check whether the positive effect in the manufacturing firm was due to training participants being on faster earnings trajectories or being better workers, they estimated the effect of training on the worker's pretraining wage level and growth. The results showed that participants did not have significantly different levels of pretraining wages, or pretraining wage growth, than nonparticipants. They found weak or no effects of training on turnover, absenteeism, and performance awards.

Through access to company records, Krueger and Rouse determined that the total cost of training (direct expenses and release time) was approximately 4 percent of the average trainee's annual compensation. Assuming that completed job tenure is 16 years, the real discount rate is 6 percent, and the value of training depreciates by 3 percent per year, Krueger and Rouse concluded that the program paid for itself in the manufacturing company. Using their data, I calculated an internal rate of return of 9 percent if training depreciates at 3 percent per year and a rate of return of 7 percent if the depreciation rate is 5 percent.

The reliability of the estimates from the Bartel (1995) and Krueger and Rouse (1998) studies depends on whether the econometric approach was successful in eliminating the bias that may result from nonrandom assignment to training programs in this company. Another limitation of these studies is that the impact of training is only estimated during the first year after training, with no direct observation of the rate at which the training impact actually depreciated. In addition, calculation of the company's ROI relies on the assumption that the magnitude of the wage gains is a good proxy for the magnitude of productivity gains.

Ichniowski et al. (1997) used monthly data on the productivity of 36 steel finishing lines. Unlike Huselid, they collected data on human resources management policies by conducting standardized, but open-ended, interviews with human resources managers, labor relations mangers, operations managers of the finishing lines, superintendents, line workers, and union representatives in organized lines. Supporting data were gathered from personnel files, personnel manuals, and collective-bargaining agreements. This enabled them to construct variables that measure practices in seven human resources management policy areas, one of which is training. The productivity measure that they studied is "uptime," the percentage of scheduled operating time that the line actually runs. They characterized each finishing line's system of human resources management policies as falling into one of four categories, ranging from highly innovative to highly traditional. Using a model that accounts for the role of fixed, unobserved characteristics, they estimated the impact of a change in the line's human resources management system on the change in productivity. The results show that moving from the most traditional system to the next most innovative system raises productivity by 2.5 percentage points and another 3.5 percentage points if the line moves up to the next most innovative system. However, introducing a high-training policy (where all line operators receive training) by itself had no impact on productivity.[19] In the case of one finishing line, Ichniowski et al. calculated that switching from the traditional human resources management system to a more innovative one that included work teams and extensive skills training resulted in a $27,900 increase in monthly profits.

To summarize, the econometric case study approach overcomes two of the problems of the large-sample econometric studies, namely, heterogeneous production processes and lack of cost data. Conclusions regarding the internal rate of return on company training programs depend on assumptions about skill depreciation. If the rate of depreciation is assumed to be 5 percent, the estimated internal rates of return range from 7 to 50 percent. However, it is not clear that the results from these studies are entirely free of the bias resulting from the endogeneity of the training decision. By definition, a nonexperimental approach requires a leap of faith that the instruments used by the researcher are, in fact, the correct ones or at least very highly correlated with the correct ones. As Lalonde (1986) has argued for public training programs, the nonexperimental approach is not a perfect substitute for the experimental approach.[20]

## Case Studies of Training Programs in U.S. Businesses

*Evaluation practices of U.S. businesses.* A third source of information on the employer's rate of ROI in employee training is the companies' own evaluations of their training programs. Surveys of evaluation methods

[19] This result is at odds with Bartel's (1994) study, in which the implementation of a training policy by itself had a significant productivity impact, whereas the implementation of other human resources policies in the absence of training did not raise productivity.

[20] Lalonde's (1986) study of a publicly provided training program showed that nonexperimental techniques could not replicate experimental results, although a two-step procedure in which receipt of training was explicitly modeled outperformed the one-step approach.

used by American businesses generally show, however, that most companies measure the impact of training by considering workers' reactions to the training, workers' learning from the training, or the impact of training on workers' behavior.[21] For example, a 1986 survey of Fortune 500 companies found that only 15 percent of the respondents measured change in performance on the job, and only 8 percent measured change in company operating results that were traceable to training.[22] A 1988 American Society for Training and Development (ASTD) poll of organizations that led in training evaluation found that only 20 percent evaluated in terms of training's economic effect on the organization.[23] Data from ASTD's Benchmarking Forum indicate that the number of companies performing some type of results-oriented evaluation grew from 27 percent of Benchmarking Forum companies in 1994 to 40 percent in 1995, but it is not clear how many of these firms convert the results from training into monetary terms in order to calculate an ROI on the training investment.[24] In fact, my thorough search of the human resources management literature for the time period 1987–1997 uncovered a total of only 16 case studies of companies for which the ROI from employee training programs was measured.[25] The failure of most firms to calculate the ROI on their training investments appears to be due to the perceived difficulties in quantifying training benefits, separating the influence of training on performance improvement from other factors, and gathering the data that are necessary for an ROI calculation.[26]

*The ideal case study.* The available case studies have the potential to address the deficiencies of the studies based on large samples of firms as well as the econometric case studies discussed in the preceding section. What are the characteristics of the ideal case study? First, a pre- and posttest control group design should be used. In this design, the evaluator randomly assigns employees to either a training group or a control group.

---

[21] These three categories correspond to the first three levels of Donald Kirkpatrick's four-level model of training evaluation in which the fourth level is results, such as increased sales or higher productivity. This model, reprinted in Kirkpatrick (1996), has been the guide for evaluating training in virtually all American businesses.

[22] See Phillips (1991:7).

[23] See Carnevale and Schulz (1990).

[24] See Bassi and Cheney (1996).

[25] An additional four companies estimated returns to training but did not provide sufficient cost information necessary to calculate the ROI.

[26] See the study by Lombardo (1989). One of the major problems is that training costs are rarely shown in accounting systems as a separate cost category, instead being embedded in selling, general and administrative expenses, or cost of goods sold.

Data are gathered from both groups through pre- and posttests. If, on completion of the program, the trained group shows greater posttest performance gains than the control group, the gains are assumed to be due to the training. An alternative approach, which, if conducted properly, has the potential to produce unbiased estimates, is a time-series design in which the training group serves as its own control group. Performance is measured repeatedly (e.g., monthly) before and after the training program. If the measures show improvement after training, the result can be assumed to be due to the training as long as there are no other contemporaneous events that may have influenced productivity. The second attribute of the ideal case study is that the evaluation of the training program should use actual measures of individual worker's productivity obtained from the company's database. This presumes that the company can define and measure an employee's performance and distinguish in a specific way between top performers and weak performers. Ideally, these performance measures are available in an existing company database rather than being collected for the express purpose of evaluating the training program. Third, the company should track the productivity measures over a long enough time period to obtain a reliable measure of the rate at which worker skills depreciate.

*Lessons from the case studies.* Table 3 provides a summary of the 16 cases for which the ROI in training was calculated. For each company, the evaluation design, the employee group, the type of training, the performance measure, and the estimated ROIs are shown. Only 3 of the companies used the preferred experimental design (one company, Southeastern Bank, used a quasi-experimental design in which a post hoc control group was constructed on the basis of observed characteristics.[27]) The predominant evaluation design is the single-group pre- and posttest in which the trainees are tested before and after the training program. In order to conclude that the observed performance gains are due to the training, it is essential that with this design the evaluator try to account for other factors that may be responsible for the performance change; unfortunately, this was not done successfully in any of the cases that use this design.

The main result that stands out from Table 3 is that the estimated ROIs are extremely high, ranging from 100 to 5900 percent, when compared with the ROIs cited in the econometric case study literature (7–50 percent). While this may reflect the fact that companies do not want weak

[27] In this design, the estimates of the impact of training will be biased if unobservables play a role in assignment to training.

TABLE 3

DESCRIPTION OF CASE STUDIES

| Company | Employee Group | Type of training | Performance measure | ROI | Comments |
|---|---|---|---|---|---|
| *I. Control group design* | | | | | |
| 1. Garrett Engine | Maintenance teams | Team building | Equipment downtime converted to $ | 125% | Only studied performance for 4 weeks. Did tasks of control group differ? |
| 2. Information Services, Inc. | Professional and support staff | Interpersonal skills | Subjective behaviors converted to $ values | 336% | Subjectivity. Behavior changes measured for 1 month only. |
| 3. Financial Services Co. | District sales managers | Selection | Reduction in turnover of branch manager trainees converted to $ value | 2140% | Exaggerated the cost of turnover |
| 4. Southeastern Bank | Managers and tellers | Supervisory skills | Supervisors rate performance 6 months after training | 370% (5 years) | Post hoc control group Subjectivity |
| *II. Single-group pre- and posttest* | | | | | |
| 5. U.S. government | New supervisors | Supervisory skills | Managers rate skills 1 week before and 6 weeks after training and estimate value of skills | 150% | Subjective valuation of skills. Will changes continue after 6 weeks? Other factors may play a role |
| 6. Hughes Aircraft | Entry-level nonmanagement business jobs | Time management Reducing rework | Supervisors rate performance before and 6 weeks after training, Std. dev. = 40% of salary | 3000% | Extrapolates result for 20 to 300 employees. Subjective evaluation of performance |
| 7. Wood Panel Plant | Supervisors | Monitoring and rewarding performance | Dollar value of product defects and accidents before and 1 yr after training | 680% | No consideration of other factors that may have changed during the year |
| 8. Midwest Banking | Consumer loan officers | Consumer lending | Average monthly loan volume compared for 6 months before and 6 months after training. Valued at net profit per loan. | 1988% | Selection bias due to best prospects being sent to course. Trainees told their performance would be monitored. But did control for some other factors. |

| | | | | | |
|---|---|---|---|---|---|
| 9. Multi-Marques | Supervisors | Time management | Self-report of changes in time use, converted to $ value using salary data | 215% | Subjective self-evaluations. Extrapolates results from 3 months to 1 year. No evidence that time spent more productively. No consideration of other factors. |
| 10. Coca-Cola Bottling company of San Antonio | Supervisors | Motivation and perform. appraisal | Three months after training, trainees report increased sales, reduced waste and absenteeism | 1447% | Highly subjective. Trainees identify $ value of training. No consideration of other factors. |
| 11. Vulcan Materials | First-line supervisors | Supervisory skills | Changes in production worker turnover before and 6 months after training | 400% | Managers of the supervisors estimated % of improvement due to training. |
| 12. Yellow Freight System | Managers | Performance Appraisal | Trainees self-report $ value of productivity gains due to training | 1115% | Highly subjective. No consideration of impact of new competitive environment. |
| *III. Time series* | | | | | |
| 13. International Oil Co. | Dispatchers | Customer service | Tracked pullout costs and customer complaint costs for 11 months before and 11 months after training | 501% | Used company database to track costs. No consideration of other factors that may impact performance. |
| 14. Magnavox Electronic Systems | Entry-level assemblers | Literacy skills | Tracked average monthly efficiency for 5 months before and 6 months after training. Converted to labor cost savings | 741% | Productivity gain occurred only in first month after training. Extrapolated 4-month average to a full year. Ignored role of other factors. |
| 15. Arthur Andersen & Co. | Tax professionals | State and local taxation | Tracked fees and chargeable hours for 12 months before and 12 months after training | 100% payback in one year | Compared changes in revenue of participating and non-partic. offices to reject role of other factors. But participants already more experienced. |
| 16. CIGNA | Managers | Mgmt. development | Manager develops productivity measure that is tracked for 3 months before and 4 months after training | 5900% | Used actual work-unit records to document performance. Trainees knew their performance would be measured after 4 months. |

SOURCES: Case 1, Pine and Tingley (1993); case 2, Phillips (1994); case 3, Phillips (1994); case 4, Mathieu and Leonard (1987); case 5, Phillips (1994); case 6, Merlo (1988); case 7, Robinson and Robinson (1989); case 8, Phillips (1994); case 9, Phillips (1994); case 10, Phillips (1994); case 11, Carnevale and Schulz (1990); case 12, Phillips (1994); case 13, Phillips (1994); case 14, Phillips (1994); case 15, Phillips (1994); case 16, Pasquet et al. (1987).

results publicized, it is probably due to a number of methodologic flaws that are observed repeatedly in the case studies. In addition to ignoring the role of factors other than training that could be responsible for performance changes, most of these cases, as noted in Table 3, suffer from at least one of the following problems: (1) using supervisors' subjective evaluations of trainees' performance levels, (2) using self-reports from the trainees about the productivity gains from training, (3) monitoring the impact of training for a relatively short time after the training is completed or ignoring the fact that the posttraining productivity gain disappeared after 1 month, (4) extrapolating findings based on a small sample of trainees to a large group of employees, (5) selecting the best employees for the training program, (6) informing the trainees that their performance will be monitored after the training program, and (7) ignoring the impact of operating in a new environment.

Two of the case studies, Garrett Engine and International Oil, do not suffer from these methodologic flaws and come close to matching the characteristics of the ideal case study described earlier. In this section, the methodologic approaches and results obtained from these two cases are discussed.

The Garrett Engine case, discussed in Pine and Tingley (1993), comes from the division of Allied Signal that manufactures jet engines. The company was concerned about the downtime of its equipment and decided to use a 2-day team-building training program for the maintenance teams that repair the equipment. Each team consisted of a manager and hourly employees such as electricians, plumbers, and mechanics. Four similar maintenance teams were identified. Two teams were randomly assigned to the experimental group and two to the control group.

Prior to the training, all teams were measured in terms of their equipment downtime, job response time, and job completion time. *Response time* is the time it takes for a team to respond to a call for service. *Completion time* is the time required to complete a job. The teams were measured on all three dimensions for 4 weeks after the training. The company found that prior to the team-building course, the people in the experimental group were slower to respond to job requests than those in the control group. After the program, the experimental group responded more quickly, whereas the control group stayed about the same. Similarly, in the case of job completion time, prior to training, the experimental group performed worse than the control group. After training, the experimental group significantly reduced its completion time, whereas the control group's time remained about the same. As a result of improvements in response time and completion time, total downtime for the experimental group fell from 18.4 to 15.8 hours; the control group's downtime stayed constant at about 16 hours.

The maintenance department had established a "burden rate," or the cost of machine downtime, prior to this study. Using this figure, the downtime cost was calculated for the experimental and control groups before and after the training. Downtime cost was $1156 per job for the experimental group after training and $1211 per job for the control group after training. This translates into a $55 savings per job that is attributable to the training.[28] The company then used the cost savings estimate to calculate the ROI on the training program. This required making an assumption about the length of time that the effect of the training would last. Using a very conservative assumption of 4 work weeks, the resulting ROI was 125 percent.

In general, the approach taken by Garrett Engine is a good one. By randomly assigning similar teams to the experimental and control groups and observing the results over a fairly short time period, the company can be reasonably confident in concluding that the performance gains observed over the 4 weeks were due to the training. The evaluators noted that in order to conclude that the training effect would last beyond the 4 weeks, they would need to monitor performance of both groups for a longer time period. An important issue of concern, however, regarding this case is that the company only studied four teams and conclusions regarding statistical significance cannot be made with so few observations.

The International Oil case, described in Phillips (1994), uses a time-series design to evaluate a training program for dispatchers at International Oil's central dispatch in Los Angeles. The training program was introduced in response to an increase in delivery costs from order errors as well as customer complaints about these errors. The training program was evaluated by tracking data on pullouts (when a dealer cannot take the full load ordered and the truck has to "pull out" and return to the terminal to adjust the product mix for the next station on the schedule), dealer complaints, and dispatcher absenteeism, before and after training. The data were tracked for 11 months prior to the training program and 11 months after. The company had already calculated the cost of a pullout. Using this information, it was able to calculate the difference between pullout costs before and after training. By using information on managers' salaries and the average amount of time it took a manager to resolve a customer complaint, the company estimated the cost of a dealer complaint. This was

[28] A less conservative estimate of the cost savings would be to measure the experimental group's performance before and after training; this resulted in $185 savings per job.

used to measure the cost savings from the reduction in dealer complaints that occurred in the 11 months after the training. Finally, the value of the reduction in dispatcher absenteeism was quantified by using data on dispatchers' salaries. Combining these savings produced the total benefit for the program. The cost of the program was underestimated because data on the trainer's salary and the cost of the training facility were unavailable; in addition, the company did not account for the lost productivity of the trainees while they were attending the training program. Hence the reported ROI, 501 percent, based on the company's reported costs of \$60,000, is likely to be an overestimate. Using data reported in the case, I adjusted this cost figure to reflect the excluded cost components and recalculated the ROI. Since the training program ran for a total of 3 weeks, with each of the 11 participants attending four 1-hour sessions, it seems unlikely that even with these additional cost components, the total training cost would more than double. A total cost of \$120,000 results in an ROI of 200 percent.

This is an interesting case because it demonstrates how data from company records can be used to successfully quantify the outcomes of a training program. Since the company had a history of tracking pullouts and dealer complaints, it could tap into this database and monitor changes over fairly long periods before and after the training. The weakness in this case is that there was no attempt to consider the fact that the dispatcher's efficiency depends on both the behavior of the truck drivers and the ability of the dealers to accurately measure the amount of gasoline in their tanks. It is possible that the reduction in pullouts occurred because there were more experienced truck drivers and/or dealers in the later time period than before the training, although it seems unlikely that the composition of these two groups changed significantly during the time period under study.

Although the vast majority of the case studies in the literature contain serious methodologic flaws, the two well-conceived case studies reviewed here calculated ROIs in the range of 100 to 200 percent. These estimates are higher than those reported in the econometric case studies reviewed earlier.[29] While it is possible that the different results are due to the unique situations of specific firms,[30] it is equally likely that the stronger results

[29] It is worth noting that experimental estimates of the impact of government training programs (i.e., JTPA) on the wages of the economically disadvantaged have produced ROIs of 74 percent for adult men and 41 percent for adult women. See Friedlander et al. (1997).

[30] The ROI in training that a firm could expect to receive is likely to depend on its competitive strategy, its production processes, and/or its other human resources practices. Hence the ROIs achieved by these two firms should not be interpreted as the ROI that other firms in different situations may receive.

reflect better measurement of training's impact resulting from the use of more appropriate measures of productivity. Both the econometric case studies used wages as a proxy for the worker's productivity, whereas the two company-sponsored case studies used actual productivity measures.

## Conclusion

Knowledge of the rate of return on an employer's investment in employee training provides guidance to firms on their human capital investment decisions and can aid government policymakers in decisions regarding subsidies of private investment. The purpose of this article was to review the literature on the employer's ROI in employee training in order to assess the usefulness of the reported estimates and to suggest directions for future research to improve the accuracy of the measured rates of return.

Three components of the literature were reviewed. The first, which uses large samples of firm-level or establishment-level data collected through mail or phone surveys, was found to provide little guidance on the rate of return because training cost data were typically unavailable, diverse production processes may not have been modeled properly, and the bias from the endogeneity of the training decision may not have been entirely eliminated. The second component of the literature, in which an econometric framework is applied to data from one or two companies to conduct an "econometric case study," addresses the first two limitations of the large-sample studies. The estimated rates of return from this literature depend on the assumption regarding the skill depreciation rate. Assuming that skills depreciate 5 percent per year, the estimated rates of return range from 7 to 50 percent. The third component of the literature, company-sponsored case studies, has the potential to address the deficiences of the studies based on large samples of firms as well as the econometric case studies. Unfortunately, few companies calculate the ROI in employee training, and among those which do, the vast majority use faulty methodologies that preclude relying on their results. The two well-conceived case studies reviewed in this article report ROIs in the range of 100 to 200 percent. To the extent that these results occur because of the ability to eliminate endogoneity bias as well as the use of more accurate measures of productivity than were available in the econometric case studies, these findings indicate that the employer's ROI in training may be much higher than previously believed.

Of course, it is not appropriate to generalize on the basis of two case studies. As documented earlier, most American businesses do not evaluate the business results of their training programs. A sound ROI analysis requires extensive data collection on numerous measures for many employees at multiple points in time. Many companies are not equipped to undertake such an effort for the express purpose of evaluating a training program. Some companies already maintain detailed records on employee performance and employee characteristics, and such companies are best suited for conducting an ROI analysis. Where possible, researchers should be encouraged to gain access to such databases and to supplement them with data-gathering efforts to collect information on variables needed to isolate the effect of training.[31] In addition, government resources can be used to encourage firms to conduct ROI analyses of their training programs. If lack of information on the ROI is one cause of possible underinvestment in employee training, helping firms to measure the returns on their training investments could help resolve the underinvestment problem.

## References


Barron, John M., Mark C. Berger, and Dan A. Black. 1993. "Do Workers Pay for On-the-Job Training?" Working Paper No. E-169-93, Lexington, KY: University of Kentucky.

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

Bartel, Ann P. 1994. "Productivity Gains from the Implementation of Employee Training Programs." *Industrial Relations* 33(October):411–25.

_____. 1995. "Training, Wage Growth and Job Performance: Evidence from a Company Database." *Journal of Labor Economics* 13(July):401–25.

Bassi, Laurie J., and Scott Cheney. 1996. *Restructuring: Results from the 1996 Benchmarking Forum*. Alexandria, VA: American Society for Training and Development.

Bishop, John H. 1991. "On-the-Job Training of New Hires." In *Market Failure in Training?* edited by David Stern and Jozef M. M. Ritzen, pp. 61–98. New York: Springer-Verlag.

_____. 1997. "What We Know About Employer-Provided Training: A Review of the Literature." In *Research in Labor Economics*, Vol. 16, edited by Solomon Polacheck, pp.19–87, Greenwich, CT: JAI Press.

Black, Sandra, and Lisa Lynch. 1996a. "Human Capital Investments and Productivity." *American Economic Review* 86(May):263–7.

_____ and _____. 1996b. *How to Compete: The Impact of Workplace Practices and Information Technology on Productivity*. Washington: U.S. Department of Labor.

Blomberg, Robert, Elizabeth Levy, and Ailene Anderson. 1988. "Assessing the Value of Employee Training." *Health Care Management Review* 13(1):63–70.

Carnevale, Anthony, and Eric Schulz. 1990. "Return on Investment: Accounting for Training." *Training and Development Journal* (July):S1–32.

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35(December):1809–55.


[31] Ideally, analysis of the ROI in training should be conducted for different types of training in order to determine which types create the most value for a firm.

Goldstein, Irwin L. 1992. "Training in Work Organizations." In *Handbook of Industrial and Organizational Psychology*, edited by Marvin D. Dunnette and Leaetta M. Hough, pp. 507–619. Palo Alto, CA: Consulting Psychologists Press.

Holzer, Harry, Richard Block, Marcus Cheatham, and Jack Knott. 1993. "Are Training Subsidies for Firms Effective? The Michigan Experience." *Industrial and Labor Relations Review* 46(July):625–36.

Huselid, Mark. 1995. "The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance." *Academy of Management Journal* 38(June):636–72.

_____ and Brian E. Becker. 1996. "Methodological Issues in Cross-Sectional and Panel Estimates of the Human Resource-Firm Performance Link." *Industrial Relations* 35(July):400–22.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. 1997. "The Impact of Human Resource Management Practices on Productivity." *American Economic Review* 87(June): 291–313.

Kirkpatrick, Donald. 1996. "Revisiting Kirkpatrick's Four-Level Model." *Training and Development* (January):54–9.

Krueger, Alan, and Cecilia Rouse. 1998. "The Impact of Workplace Education on Earnings,Turnover and Job Performance." *Journal of Labor Economics* 16(January):61–94.

Lakewood Publications. 1997. "1997 Industry Report." *Training* (October).

Lalonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76(September):604–20.

Loewenstein, Mark A., and James R. Spletzer. 1998. "Dividing the Costs and Returns to General Training." *Journal of Labor Economics* 16(January):142–71.

Lombardo, Cynthia A. 1989. "Do the Benefits of Training Justify the Costs?" *Training and Development Journal* (December):60–4.

Mathieu, John, and Russell Leonard. 1987. "Applying Utility Concepts to a Training Program in Supervisory Skills." *Academy of Management Journal* 30(April):316–35.

Merlo, Nicholas S. 1988. "Subjective ROI." *Training and Development Journal* (November): 63–6.

Milgrom, Paul, and John Roberts. 1995. "Complementarities and Fit: Strategy, Structure and Organizational Change." *Journal of Accounting and Economics* 19(April):179–208.

Paquet, Basil, Elizabeth Kasl, Laurence Weinstein, and William Waite. 1987. "The Bottom Line." *Training and Development Journal* (May):27–33.

Phillips, Jack J. 1991. *Handbook of Training Evaluation and Measurement Methods*. Houston, TX: Gulf Publishing Company.

_____. ed. 1994. *Measuring Return on Investment*, Vol. 1. Alexandria, VA: American Society for Training and Development.

Pine, Judith, and Judith C. Tingley. 1993. "ROI of Soft-Skills Training." *Training* (February): 55–60.

Robinson, Dana Gaines, and Jim Robinson. 1989. *Training for Impact*. San Francisco, CA: Jossey-Bass.

Tan, Hong W., and Geeta Batra. 1995. "Enterprise Training in Developing Countries: Incidence, Productivity Effects and Policy Implications." Unpublished paper, The World Bank.