

Princeton University
Industrial Relations Section
Working Paper #60
December 1974

The Effect of Manpower Training on Earnings: Preliminary Results

Orley Ashenfelter*
Princeton University

* This paper will be presented at a joint session of the American Economic Association and the Industrial Relations Research Association in San Francisco, December 28-29, 1974.

The Effect of Manpower Training on Earnings: Preliminary Results

Orley Ashenfelter^{1/}
Princeton University

The stakes involved in a judgment on the economic success of post-schooling classroom training are large. On the one hand, through fiscal year 1973 we had invested some \$2.5 billion in this social program through federal obligations for institutional training under the Manpower Development and Training Act (MDTA). As citizens we should like to know how well this investment paid off so future decisions can be altered in the light of the success or failure of decisions made in the past. On the other hand, post-school classroom training is almost like a laboratory test of our ability to augment the human capital of certain workers. As academics we should like to know what this experiment can tell us about the theory of human capital and the value of this theory for purposes of prediction and measurement.

^{1/}This paper is a progress report on a longer-term empirical study of the effects of the Manpower Development and Training Act programs on trainee earnings. Much of it was started in 1972 when I was Director, Office of Evaluation, U.S. Department of Labor. I owe a debt to Laurence Silberman, then Under Secretary of Labor, and Michael Moskow, then Assistant Secretary for Policy, Evaluation, and Research, for the opportunity to pursue this line of inquiry. Among many others in the Department of Labor, I am heavily indebted to James Blum for many helpful discussions and to David Farber for introducing me to the data system I have so heavily used. I also must thank Sherwin Rosen of the University of Rochester and Ernst Stromsdorfer, then of Indiana University and now of the Department of Labor, for serving as Chairman and Consultant, respectively, of the Panel on Manpower Training Evaluation of the National Academy of Sciences. The final report of the Panel, The Use of Social Security Earnings Data for Assessing the Impact of Manpower Training Programs (Washington, D.C.: National Academy of Sciences, January 1974), constitutes a lucid and independent analysis of the data base on which this paper is based. Finally, D. Alton Smith masterfully performed most of the computing work underlying this paper, and assisted in numerous other ways. Financial support for this stage of the work has been provided by the Office of Evaluation, ASPER, U.S. Department of Labor. Needless to say, none of those mentioned above is responsible for the views expressed in this paper.

Regardless of the stated objective, manpower training programs are successful only if they increase trainee earnings above what they otherwise would have been. The stated objectives of manpower training programs fall into three categories: (1) The reduction of inflation by the provision of more skilled workers to alleviate shortages, (2) the reduction of the unemployment of certain groups, and (3) the reduction of poverty by increases in the skills of certain groups. Since the first objective requires the alleviation of shortages by training more highly skilled workers, training should increase the earnings of those whose skills were increased. The concern for reducing unemployment, which is the second objective, is derived from a concern for the decreased earnings of unemployed workers, and if trainees subsequently suffer less unemployment their earnings should be higher. Finally, training programs are designed to reduce poverty by increasing the earnings of low income workers. It follows from this discussion that an analysis of the success of training programs is inherently a quantitative analysis of how trainee earnings are changed from what they otherwise would have been. The purpose of this paper is first to analyze the advantages and disadvantages of performing such an analysis using longitudinal data on trainees that is obtained from Social Security Administration records, and then to report on what I have learned from the application of this method to one cohort of trainees who participated in institutional (classroom) training under the MDTA.

I.

Although there have been many analyses of the effect of post-school manpower training on earnings it is generally agreed that very little is reliably known about the actual effect of these programs.^{2/} There are three main problems that plague all efforts to analyze these programs: (1) The large sample size required to obtain reliable estimates of program effects, (2) the great expense of interviewing trainees over several years, and (3) the difficulty of obtaining an adequate group against which to compare trainees.

To illustrate the first point, suppose that the second and third problems were resolved and that we wished to design a study of the impact of training on earnings. A simple, but adequate, design would be to take a group of applicants for the program and randomly assign only some proportion π to the program, reserving the remainder, $1-\pi$, as a comparison group. The program's average effect on earnings would then be estimated by the difference after training between the average earnings of trainees and the comparison group. Of course, this difference would only be an estimate of the program's true average effect because it would be based on only a sample of the trainees. We are naturally interested in how reliable this estimate would be. One measure of reliability is the

^{2/}Jon Goldstein in The Effectiveness of Manpower Training Programs: A Review of Research on the Impact of the Poor (Washington: U.S. Government Printing Office, 1972), p. 14 puts it this way: "The robust expenditures for research and evaluation of training programs (\$179.4 million from fiscal 1962 through 1972) are a disturbing contrast to the anemic set of conclusive and reliable findings."

standard deviation of the estimated program effect, since, after all, we should expect the interval defined by the estimated program effect plus or minus two standard deviations to contain the true program effect in some 95 out of 100 designs. For this simple design the standard deviation of the program effect, σ_t , is given by^{3/}

$$(1) \quad \sigma_t = \sigma_y / [N(\pi)(1-\pi)]^{1/2},$$

where N is the overall sample size, and σ_y is the standard deviation of earnings in the potential trainee population. σ_t clearly depends inversely on our choice for N , mirroring our intuition that large samples should provide more reliable information.

How small should we require σ_t to be? A formal answer to this question is very difficult, but surely the answer will depend on how large we think the true program effect is likely to be. In 1964, total federal costs were around \$1,800 per MDTA institutional trainee. If the rate of return on this investment were .10, and if the time stream of returns involved a one-time increase in trainee earnings, we would expect a training effect of \$180. Doubling our expected rate of return would double this number. Given a training effect of around 200 (1964) dollars, it is clear that a 95 percent confidence interval could be expected to include zero unless σ_t were less than \$100. So wide a confidence interval would then preclude even the judgment that the training effect was significantly different from zero and probably would not be of much help in decision-making. Therefore, to continue the example in concrete terms let us

^{3/}This is just one form for the standard deviation of the difference between two sample means, and can be found in any elementary statistics book.

suppose that we will require at the very least that $\sigma_t \leq 100$. In terms of formula (1) this requires that

$$(2) \quad \sigma_t = \sigma_y / [N(\pi)(1-\pi)]^{1/2} \leq 100.$$

Inverting the inequality in (2) shows that this requires

$$(3) \quad N \geq (\sigma_y/100)^2 / \pi(1-\pi).$$

To see how large the sample size, N , must be then requires only estimates of σ_y and π . Clearly the required value of N is lowest for $\pi = .5$, an equal split of the sample between trainees and controls, and so we might as well use this value. In the data for 1964 that I discuss below σ_y is around \$2,000 for white males, though it is slightly lower for other race/sex groups. Putting these numbers together in formula (3) gives $N \geq (2,000/100)^2 / .5(1-.5) = 1600$.^{4/} Although very much larger than the completed sample size in many studies, $N = 1600$ does not seem unreasonable until one recognizes that it would be necessary to have at least 1600 sample points for every separate race/sex or other group that was to be analyzed. With two race groups and two sex groups, for example, it would be necessary to have 6400 sample points, which is considerably larger than any sample actually drawn.

This is hardly the end of the story, however, because the difficulty and expense of interviewing trainees and controls over several years still

^{4/}I should emphasize that these calculating are meant only as examples. In practice, a more complicated regression design would surely be used. To the best of my knowledge, the first formal consideration of these issues in the context of the actual design of a study to analyze a Department of Labor training program was taken up by George Johnson and H.M. Pitcher in "On the Necessary Sample Size for an Impact Evaluation of the WIN 2 Program," unpublished paper, November 1973. Their analysis considers the more complicated case.

remains. Suppose, for example, that it costs R dollars to complete a single interview and that the probability of completing an additional interview is λ . This latter assumption means that the success rate for interviews is a constant λ and independent of the number of consecutive interviews. In this case the cost of a completed sample point that requires k interviews is just

$$(4) \quad C = R \sum_{r=0}^{k-1} \lambda^r / \lambda^{k-1} .$$

It is my impression that for many survey houses λ is around .9 or lower, and that R is perhaps \$25. This would imply that for a completed sample point that required four consecutive interviews $C = 25(4.72) = \$118$. Under these assumptions 1600 completed sample points would cost nearly \$200 thousand and 6400 sample points would cost about \$.75 million. There would, of course, be many other expenses to such an analysis, but this example should give some feeling for the very considerable financial obstacles that have been roadblocks to conventional analyses.

The most serious problem, however, is the difficulty of obtaining a reliable comparison group for the analysis. Manpower training programs have not been run as if they were experiments in the past, and there is a great deal of difficulty and resistance in changing past practices. The argument against experimentation is usually shrouded in the rhetoric of morality, implying that deliberate exclusion of any group from entrance into a training program to which they were entitled is unethical. The fact that this argument is entirely fallacious is demonstrated by two simple points. First, virtually all MDTA programs have offered

substantial stipends to participants, so that there has normally been an excess supply of applicants. This means that some eligible trainees have always been excluded from training, and must have been excluded by some criteria. In response to this point it might be argued that those applicants that are accepted into training have been chosen by program operators on the basis of the likelihood that they would benefit from the program. There is, of course, no reason why this policy could not be continued in an experimental environment so long as the basis of the selection criteria were known and measurable, in which case the selection criteria could be controlled statistically. No doubt there would be resistance to requiring the practice of using explicit and objective criteria for entrance to the program, but such a practice might be desirable on grounds of fairness even in the absence of an experimental environment.^{5/}

II.

Each MDTA trainee had created for him (or her) a trainee file upon program entrance. No doubt the original idea for these files was inspired by the analogy of trainees with school students, and their purpose was to allow the analysis of a trainee's progress through the program and into the labor market. As in the case with school students, the latter turned out to be difficult, and in practice the files do not contain useful information on the subsequent labor market status of trainees.

^{5/}The statistical issues involved are discussed by Arthur Goldberger, "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations," Discussion Paper No. 123-72, Institute for Research on Poverty, The University of Wisconsin, April 1972.

These basic trainee files make up one set of data for analysis.

At the same time, the Social Security Administration maintains information on the earnings of most American workers year by year over the last two decades or so. These data are obtained from employer records and are maintained, of course, for the purpose of determining retirement and other benefits of eligible workers. In the analysis I report below the basic data source for trainees is the combination of these Social Security records with the training program files after deletion of identifying characteristics to preserve confidentiality.

This source of data goes a long way toward resolving two of the basic problems in estimating the effect of training on trainee earnings that I have outlined above. First, there is no problem with sample sizes in these data. In principle it would be possible to use the data for all trainees. In practice, since the earnings data are available only on a calendar year basis it seems better to work with trainees entering the program during the first calendar quarter of a year so that we have an uncontaminated measure of trainee earnings covering the entire period prior to training. Table 1 contains the sample sizes for data obtained in this way for 1964 enrollees. As can be seen from the table, these are certainly large enough to give statistically reliable results on the criterion outlined above. Second, there is no additional cost to obtaining an extra year's earnings for any trainee. The only limit on the length of time that the earnings of a trainee might be followed is the number of years since leaving training. Since 1964 was the first year

Table 1

Sample Statistics for MDTA Trainees, 1st Quarter of 1964 Enrollees

Race, Sex <u>Group:</u>	Sample Size		Average Earnings in 1963	
	<u>Trainees</u>	<u>Controls</u>	<u>Trainees</u>	<u>Controls</u>
White Males	7,326	40,921	1,810.	3,108.
Black Males	2,133	6,472	1,181	1,896.
White Females	2,730	28,142	748	1,432.
Black Females	1,356	5,192	531.	936.

there was a sizeable MDIA trainee population, in this preliminary report I have concentrated on the results for white males in the 1964 cohort because it allows analysis of the effect of training on earnings for a full five years after training. I should add that the average cost of obtaining and processing a completed trainee sample point in this way has been less than \$.50. For a comparison group I have been forced to use the .1 percent Continuous Work History Sample (CWS). This is a random sample of American workers that is maintained by the Social Security Administration. The sample sizes for the comparison group are also listed in Table 1. In the case of both the trainee and comparison groups I have limited the sample to those persons between the ages of 16 and 64.

It is obvious that the comparison and trainee groups are not drawn from the same population. The best that can be done in these circumstances is to statistically control for observable differences between the two groups using measurable variables. I have chosen to do this by regressions that include as separate independent variables the five previous years' earnings and age.

I do not want to over-sell the reliability of this method of controlling for differences between the trainees and other workers. Even so, the use of longitudinal data in this way is a powerful tool for two reasons. First, there are a variety of historical factors that influence the earnings capacity of any individual worker, such as education, experience, mobility, and others. Surely the best single summary measure of this cumulative experience for a worker is his previous earnings, and it is precisely these differences in pre-training earnings that are used to control for differences between the trainees and the comparison group.

Second, it is possible to provide an internal check on the quality of the standardization by fitting the same model used to estimate differences between the trainee and control group in the period after training to the data for the period prior to training. That is, I have not only estimated the standardized (or controlled) difference between the average earnings of trainees and controls after training, but also the standardized difference before training. Needless to say, if the regression model were entirely successful we would expect the estimated pre-training differences to be negligible.

The basic results for white males are contained in Table 2 in the form of estimated regression coefficients (and standard errors) of a dummy variable that equals unity for trainees and zero for others. I have suppressed the full regression results to conserve space (there is one regression for each number reported in Table 2), but the explained variance ($R^2 \approx .7$) and other sample statistics are very impressive. Several conclusions may be drawn from this table. To begin with, it is clear that in 1963 the standardized earnings of the 1964 cohort of trainees were roughly \$300 to \$350 lower than the standardized earnings of the controls. The difference between trainees and controls in 1962, however, was very small (\$42). This means that the earnings of trainees must have fallen relative to those of the controls in the year immediately preceding training. In retrospect this finding is probably not surprising, but in so far as I am aware it has never been reported in any of the voluminous studies of this issue. What it implies is that manpower trainees in 1964 were not just workers from the lower end of the earnings distribution as shown in Table 1, but that they were workers who had special and substantial labor /
market

Table 2

Estimated Standardized Differences Between Trainee and Control Group
Earnings, White Male Institutional MDTA Trainees from the Class of 1964^{a/}

Difference Between Trainees and Controls In:	Earnings Differences are Standardized for all Years Equal or Prior to:		
	1961	1962	1963
1962	-42. (13.)		
1963	-347. (16.)	-317. (13.)	
1964	-907. (18.)	-883. (17.)	-682. (15.)
1965	139. (20.3)	161. (19.)	322. (25.)
1966	150. (26.)	176. (26.)	342. (25.)
1967	138. (28.)	163. (28.)	322. (18.)
1968	-7. (33.)	19. (33.)	168. (32.)
1969	62. (36.)	88. (35.)	225. (35.)

^{a/} Estimated standard errors of estimated regression coefficients in parentheses.

Conclusion

It is my general impression that just as manpower training was considered the panacea of the 1960's for resolving the problems of chronically unemployed workers, it is now widely believed that these training programs were failures. As I have suggested in this paper, neither judgment about manpower training programs has been supported by any careful empirical analysis. One is led to wonder how such wild swings in public opinion are generated in the face of such staggering ignorance about the actual as opposed to the intended effects of a program. As I have also suggested, the earnings of the 1964 cohort of MDTA trainees do seem to have been raised by training above what they otherwise would have been, though the absence of an experimental design for this study puts a considerable range of uncertainty into the estimates. Clearly further work with the data system described here would be desirable, and I intend to report further results in the future. Still, there will never be a substitute for a carefully designed study using experimental methods, and there is no reason why this could not still be carried out. But perhaps it is too late for such a study since the panacea peddlers have lost interest in training programs and have now turned to the notion of a massive public employment program as the panacea of the 1970's.