A Meta-Analysis of Government Sponsored Training Programs

by

David H. Greenberg
University of Maryland, Baltimore County

Charles Michalopoulos
MDRC

Philip K. Robins
University of Miami

September 2001

The research reported in this study was supported by a grant from the Rockefeller Foundation to the Maryland Institute for Policy Analysis and Research, University of Maryland, Baltimore County. All opinions expressed are those of the authors and do not necessarily reflect the views of the Rockefeller Foundation or any of its employees or sponsoring organizations. We would like to thank Howard Bloom and Marvin Mandell for valuable advice during the course of this research.

EXECUTIVE SUMMARY

This report synthesizes the effects of 31 evaluations of 15 voluntary government training programs for the disadvantaged that operated in the United States between 1964 and 1998.[1] The 15 programs were quite diverse. They used different types of training, including structured job search, remedial education, classroom vocational or skills training, on-the-job training (OJT), and subsidized employment in the public or private sector. Some of the programs were national in scope, but others were limited to specific geographical areas. Some were based on random assignment, but other evaluation designs were also used.

This study uses *meta-analysis*, a technique that statistically synthesizes diverse studies. Meta-analysis provides information that cannot be found in individual evaluations or in other types of syntheses. At best, individual evaluations indicate whether particular programs “worked,” but they provide very limited information on why they were effective, for whom they were most effective, and the economic conditions under which they were most effective. Syntheses that have not used meta-analysis have focused more on the overall effectiveness of the programs than on the factors that cause one program to be more effective than another.[2] The objective of the research described in the report is to explore which training types are most effective, the sorts of people for whom they are most effective, and the circumstances under which they are most effective.

The main findings of the study include the following.

- **The programs were most effective for women, only modestly effective for men, and generally ineffective for youth.** The average effect of the programs was more than $1,400 for women (in 1999 dollars), about $300 for men, and close to zero for youth. Even for women,

---

[1]Government-funded training programs can be divided into voluntary and mandatory programs. Voluntary programs for the disadvantaged are for individuals who meet certain criteria relating to need or economic disadvantageness and are seeking jobs or to upgrade their skills. Mandatory programs, which are mainly aimed at welfare recipients, require participation in exchange for benefits. As discussed in Friedlander, et al. (1997), findings for these two types of programs cannot be readily compared. The present study examines only voluntary programs.

[2] For example, see LaLonde, 1995; Friedlander, et al., 1997; Heckman, et al., 1999.

where the average effect is sizable, there were few individual cases where the effects were large. The vast majority of evaluation estimates for women indicate that government-funded training programs have increased earnings by less than $2,000 a year.

- **In general, classroom skills training was found to be effective, but basic education was not.** For purposes of the meta-analysis, we classified training into one of six categories: classroom skills training, basic education, a mix of classroom skills training and basic education, OJT (on-the-job training), subsidized work, and a mix of classroom and workplace training. For women, all types of training other than basic education were about equally effective. For youth, only classroom skills training appeared to be effective. For men, it was difficult to draw firm conclusions about which types of training were effective or ineffective because the estimated effects of different training types varied considerably with demographic and economic circumstances of the trainees.

- **Among similar programs for men and women, more expensive programs were not more effective than less expensive programs, but they may have been more effective for youth**. One possible explanation for this finding is that more expensive programs were inefficiently run and did not provide better services for their participants.

- **The effects of the programs do not appear to deteriorate over time, at least in the short term**. In contrast to other research, we found little evidence that the effects of training on the earnings of adults diminished over time once participants left the programs. In fact, in the year or two after training, the effects of the programs appeared to increase for adults. However, we observed the effects of training for a maximum of five years after the training was completed.

- **Recent programs do not appear to be more effective than earlier programs**. Although the United States has more than three decades of experience in running training programs, we find no evidence that voluntary training programs for the disadvantaged have become more effective over time in increasing earnings.

- **Higher unemployment rates seem to decrease the effectiveness of training for youths.** However, unemployment rates do not appear to be related to training effects for women, and the effect for men are unclear.

- **Training seems less effective for white men and white youth than for non-white and racially mixed groups.** One possible explanation for this finding is that white workers face fewer employment barriers and, hence, can more readily find jobs on their own without the aid of training. For women, the effects of the programs were about the same for different racial groups.

A Meta-Analysis of Government Sponsored Training Programs

1. Introduction

Since the 1960s, federal and state governments have funded training programs aimed at increasing the earnings of low-income individuals who have ended their formal education. They have envisioned these programs as tools for combating unemployment and poverty and, more recently, as integral to increasing the earnings of government transfer program recipients and, thereby, decreasing their benefit payments. Evaluations of the effectiveness of government funded training programs for the disadvantaged have been accumulating for more than three decades.

The objective of the research described in this report is to increase knowledge of the types of government training programs for the disadvantaged that are most effective, the sorts of persons for whom they are most effective, and the circumstances under which they are most effective. As explained below, this will be done primarily by using a well established set of formal statistical techniques known as meta-analysis to exploit variations in findings from evaluations of training programs that have estimated the effects of these programs on earnings. There have been few previous attempts to do this. At best, individual evaluations of training programs indicate whether the particular program under investigation "works"– for example, whether the program increases the earnings of those who participated in it. Individually, the evaluations provide very limited or no information on program features that are especially effective, the types of individuals that benefit most by participating in training programs, or whether economic conditions such as the level of unemployment influence program success.

Recent summaries of government-sponsored training program focus more on the overall effectiveness of the programs (for example, see LaLonde, 1995; Friedlander, Greenberg, and Robins, 1997; Heckman, LaLonde, and Smith, 1999) than on factors that cause one program to be more effective than another. For example, Friedlander, Greenberg, and Robins (1997) conclude that, while the evidence is unclear for men, previous evaluations indicate that government training programs result in modest positive effects on the earnings of adult women (especially welfare recipients), but fail to produce positive earnings effects for youths. Other summaries of the evidence from evaluations draw similar conclusions. The summaries also document that findings for all three groups have varied considerably across different evaluations, suggesting that some training programs are more successful than others, but do not probe deeply as to the reasons why.

Hence, it is evident that government training programs are often, but not always, effective in increasing earnings. Moreover, when they are effective, they are more successful in some instances than in others. Little is known, however, about why this is the case. This is not to say that there have been no attempts to explain why the findings from evaluations of government-funded training program vary. There have been, but they usually have been somewhat ad hoc in nature and their conclusions may therefore be suspect. More specifically, they rarely used formal statistical procedures to explain differences among findings, depending instead on searching for patterns in the impacts measured for different programs that are somehow related to ways in which the programs vary from one another in other respects – for example, in terms of the characteristics of the evaluated programs, the characteristics of the participants, or labor market conditions. While conclusions drawn from this approach are often plausible and intriguing, there is no firm basis for determining whether or not they are accurate because formal statistical tests are not conducted of whether the potential explanatory variables are, in fact, related to the observed variation in the estimated program effects. Thus, any conclusions based on this approach are problematic.[1]

For policy purposes, it would be extremely useful to know whether variations in estimates of program effects on earnings are attributable to differences in the services that program participants received, differences in the economic environments in which the programs operated, variation in the characteristics of the participants themselves, or some other factor such as the methods used to obtain the

[1]Greenberg, Meyer, Michalopoulos, and Wiseman (2001) use the tools of meta-analysis to demonstrate this point more rigorously.

estimates of program earnings effects. The study described in this report uses meta-analysis, which relies on statistical techniques to synthesize research findings across evaluations, to examine this topic systematically.

The meta-analysis is based on 31 studies of 15 voluntary government training programs for the disadvantaged that operated between 1964 and 1998 in the United States.[2] The components of these programs varied, but consisted of one or more of the following: remedial education, classroom vocational or skills training, on-the-job training (OJT), and subsidized employment in the public or private sector. Most of the programs also offered structured job-search, but this was always combined with one of the other program components. Some of the evaluated training programs were national in scope, but others were limited to specific geographical areas. Some of the evaluations were based on random assignment, but other (non-experimental) evaluation designs were also used. Following the lead of most training program evaluations, adult men, adult women, and youths (both boys and girls) are analyzed separately in this study.

Specifically, this study addresses the following questions:

- Do the estimates of the effects of government funded training programs on earnings differ statistically between men and women and between youths and adults? Should these groups be analyzed separately?
- Are estimated program effects on the earnings of adult men, adult women, and youths that are averaged over all evaluations of voluntary training programs positive and statistically significant for each of these groups?
- Do more expensive training programs, as measured by cost per program participant, result in larger effects on earning than less expensive programs?
- Do effects on earnings differ among programs providing different combinations of services? For example, does classroom training appear to be more or less effective than on-the-job training?

[2]Government funded training programs can be loosely divided into voluntary programs and mandatory programs. Participants in voluntary programs for the disadvantaged are individuals who meet certain criteria relating to need or economic disadvantageness and are seeking jobs or to upgrade their skills. Mandatory programs, which are mainly aimed at welfare recipients, require participation in exchange for benefits. For reasons discussed in Friedlander, Greenberg, and Robins (1997), evaluation findings for these two types of programs cannot be readily compared. Thus, the present study examines only voluntary programs.

- Are training program effects on earnings larger or smaller for some groups (for example, white participants) than for other groups (participants from minority groups)?

- Are program effects on earnings influenced by labor market conditions, as indicated, for example, by unemployment rates or the percentage of the work force in the manufacturing sector?

- Do the effects of training programs on earnings change over time after participants leave the programs (for example, are program earnings effects smaller the third year after completing the training than during the second year)?

- Do program earnings estimates that are based on random assignment experimental evaluation designs differ from those that are based on non-experimental designs? For example, do they tend to be larger or smaller? Do they tend to vary more or less from one another?

- Have training programs in recent years been more successful in increasing the earnings of participants than older training programs? In other words, has something been learned over time about how to run more successful government funded training programs?

The remainder of this report is divided into five major sections. Section 2 presents additional information about meta-analysis and discusses how it can be used to address the questions listed above. Section 3 describes the data used in the analysis. Section 4 presents descriptive statistics on the sample of evaluations used in the meta-analysis. Section 5 describes the statistical model used in the meta-analysis. Section 6 presents the study's findings. Section 6 examines the sensitivity of the findings to extreme values of the training effects. The final section summarizes the findings, discusses limitations of the analysis, and draws some policy conclusions.

2. Meta-Analysis

As discussed above, determining why findings from evaluations of government funded training programs vary can provide important information about the types of training programs that are most effective, and the types of persons for whom and the labor market conditions under which they are most effective. Meta-analysis allows one to determine whether the variation in findings from evaluations of government training programs is statistically significant and, if it is, attempts to determine the causes of this variation – for example, whether it is due to differences in the level of program expenditures, the mix of program participants, economic conditions in the places and the time periods in which the evaluations took place, the type of services provided by the programs, or in the methodologies used in the evaluations. It can also determine whether the findings vary because of an unobserved component and whether this unobserved component is due to sampling error (the fact that the underlying evaluations are derived from samples that vary in size across evaluations) or unmeasured factors (sometimes called structural error or random effects error). Meta-analysis consists of procedures for extracting findings and other information from empirical research studies and other pertinent sources, assembling this information into a data base, and then analyzing these data using modifications of standard statistical methods. Good descriptions of meta-analysis are available in Hedges (1984, 1992), Cohen (1988), Rosenthal (1991), Hunter and Schmidt (1990), Cooper and Hedges (1994), and Lipsey and Wilson (2001).

The procedures used in meta-analysis are well developed, but until recently have not been applied to government training programs.[3] They have been, however, used extensively in numerous other areas of research including medicine, psychology, education, and criminal justice. Descriptions of recent applications in other contexts can be found in Jarrell (1990), Hedges, Laine, and Greenwald (1994), Hunt (1997), and several chapters of Cook .et al. (1992).

Previous meta-analyses have followed a number of distinct and somewhat similar steps, although the exact steps depended on the objectives of the particular study and the nature of the data being used. The specific steps followed in this study, most of which are described in some detail in the following sections of this report, are listed below:[4]:

---

[3]An approach that shares important features with meta-analysis, hierarchical linear analysis, has very recently been used to analyze a subset of mandatory welfare-to-work programs (see Bloom, Hill, and Riccio, 2001). In addition, the Centre for Research in Social Policy at Loughborough University in England is currently conducting a meta-analysis of U.S. mandatory welfare-to-work programs. Perhaps, the earliest discussion of applying meta-analysis techniques to synthesizing evaluations of government funded training programs was Greenberg and Wiseman (1990).

[4]Much of the work on the first two of these steps was conducted with the late Daniel Friedlander, as part of an extension of earlier work summarizing evaluations of government funded training programs (see Friedlander, Greenberg, and Robins, 1997). We are indebted to Daniel Friedlander for his contribution to this study.

- Reports were obtained for all the evaluations of voluntary government-funded training programs of which we were aware.

- The key information from the reports, including estimates of program effects on the earnings of participants, was extracted and incorporated into a data base for use in the statistical analysis.

- Information on environmental characteristics (e.g., unemployment rates, percentage of work force in manufacturing, and AFDC payment levels) pertaining to the area where the programs were conducted was extracted from government data sources and added to the data base.

- The Gross Domestic Product chain-type price index, which is published by the U.S. Department of Commerce, was used to adjust all the monetary figures in the data base to 1999 dollars.

- The mean values of the earnings effect estimates for the training program evaluations in the data base were calculated separately for adult males, adult females, and youths. In addition, measures were constructed that indicate the extent to which program effect estimates vary across studies (the standard deviation of the earnings effect estimates). It is standard practice in meta-analysis to use the inverse of the variance of each of the earnings effect estimates as part of a weight in computing mean effects. From the earnings effect estimates and their standard errors, we also determined the level of statistical significance of each of the earnings effect estimates, as well as the statistical significance of the mean effects.[5] This is useful because some individual evaluations of training programs have found positive effects on earnings and others have found negative effects and both the positive and negative effects have sometimes been statistically insignificant (see Friedlander, Greenberg, and Robins, 1997). Hence, it is important to determine whether a statistically significant effect exists after pooling across a large number of studies, and, if it does, whether this effect is positive or negative.

- The key step in this study was to use our data base to estimate a weighted multivariate regression model that allowed us to examine the influence of individual covariates on the earnings estimates, while holding other variables constant (see below for a precise description of the regression model used).

As indicated in the last two steps listed above, statistical computations in meta-analysis rely on weights that are usually based on the variance of the estimates of program effects. The reason for doing this is intuitive. In program evaluations, estimates of program effects are obtained by using samples from the program's target population. A subset of persons from this population who participate in the program are compared to a subset of persons from the same population who do not participate. As a result of sampling from the target population, the program effect estimates are subject to sampling error. The

---

[5]In some cases, the studies only reported whether estimated program effects exceeded the 1-, 5-, or 10-percent level of statistical significance and not the standard errors of the estimated program effects. In these cases, as discussed later, we imputed values of standard errors using the effect estimate and its level of statistical significance.

variance of an estimated effect (which typically becomes smaller as the size of the underlying sample increases) indicates the size of this sampling error. In general, a smaller variance implies a smaller sampling error and, hence, that an effect estimate is statistically more reliable. Because all estimates of program effects are not equally reliable, they should not be treated the same. By using the inverse of the variance of the effect estimates as a weight, estimates that are obtained from larger samples and, therefore, are more reliable contribute more to various statistical analyses than estimates that are less reliable.[6]

Additionally, meta-analysis accounts for errors due to unmeasured factors. It is virtually impossible to measure all of the underlying factors contributing to variation in program effects across studies. Meta-analysis hypothesizes a "random effects" error in addition to a "sampling error" and, as discussed later, offers a number of alternative procedures to incorporate the random effects error into the estimation of mean effects.

3. Data

a. The Sample of Studies

As previously indicated, we have conducted an exhaustive review of every evaluation of voluntary government training programs conducted since 1964 that we could locate. In many meta-analyses, there is a concern with "publication bias," the possibility of overlooking pertinent studies (especially those with statistically insignificant effects) that have not been published in refereed journals or commercially published books. For at least two reasons, this is less of a concern for a meta-analysis of evaluations government funded training programs, even though many of the relevant studies have not been published in journals. First, most of the studies have been funded, and hence publicized, by government agencies. Second, many of the evaluations can be initially identified through citations in

[6]Using the inverse of the variance of the program effect estimates as weights in weighted least squares regressions is also an appropriate adjustment for heteroscedasticity, which may occur when different observations in regressions are based on samples of different sizes.

various review articles.

A number of studies of voluntary training programs that we uncovered, especially those dating from the 1960s and early 1970s, are nonetheless excluded from our meta-analysis. Many of these studies do not provide estimates of program effects on earnings. In some other studies that do provide earnings effect estimates, the estimates are unreliable; for example, they are based on simple before and after comparisons or rely on comparisons that are likely to be highly biased, such as comparisons between a program group of persons completing a program and a control group of program "no shows" or dropouts. In addition, one study (Finifter, 1987) did not provide the information needed to determine the variance of the earnings effect estimates, which, as noted earlier, is necessary for deriving appropriate weights.

In all, usable information from 31 studies of 15 voluntary training programs were obtained for the meta-analysis data base. Two of these studies (Kiefer, 1979 and Gay and Borus, 1980) each evaluated four separate programs, while the rest evaluated only a single program. Additionally, most of the studies separately estimated program effects on earnings for several different subgroups and for several different years after participants had left the programs, and several of the studies provide separate effect estimates for program participants at different locations. As a result, the total number of program effect estimates is much larger than 31.

Table 1 provides summary information on the training programs and evaluations included in the meta-analysis. The table classifies each program by whether it was national in scope or a demonstration program in a particular geographic area, the years over which the program operated (not necessarily the years covered by the evaluations), the demographic groups targeted, the principal program activities,[7] the major evaluation studies of the program, and the evaluation method used (experimental or non-experimental).

b. The Measure of Program Effects

[7]Job search assistance is not explicitly listed as a separate program activity because almost all the programs provided such assistance, either formally or informally or both.

Table 1
Training Program Evaluation Studies
Voluntary Programs

Program	Scope of Program	Years of Operation	Target Group	Main Activities[c]	Evaluation Study	Method of Evaluation
MDTA	NAT	1962-1973	Disadvantaged adults and youth	CT, OJT	Ashenfelter (1978), Cooley et al. (1979), Kiefer (1978, 1979), Gay and Borus (1980), Bloom (1984)	NXL
NYC	NAT	1964-1973	Disadvantaged youth	CT, PWE	Kiefer (1979), Gay and Borus (1980)	NXL
JOBS68	DEM[a]	1967-1973	Disadvantaged adults	OJT	Kiefer (1979), Gay and Borus (1980)	NXL
JC	NAT	1964-present	Disadvantaged youth	CT, PWE	Cain (1968), Gay and Borus (1980), Kiefer (1979), Mallar et al. (1982) Schochet, Burghardt, and Glazerman (2000)	NXL XL
S-D	DEM	1971-1978	Low income adults	CT	Dickinson and West (1983)	XL
CETA	NAT	1973-1983	Disadvantaged adults and youth	CT, OJT, PWE, PSE	Westat (1984), Ashenfelter and Card (1985), Bassi (1983, 1984), Bloom (1987), Bryant and Rupp (1987), Dickinson et al. (1984, 1986, 1987a, 1987b)	NXL
SW	DEM	1975-1978	Long-term AFDC recipients, ex-addicts, ex-offenders, high school dropouts	PWE with training	Hollister, Kemper, and Maynard (1984), Couch (1992)	XL
HHA	DEM	1983-1986	AFDC recipients	PWE with training	Bell and Orr (1994)	XL
TOPS	DEM	1983-1986	AFDC recipients	OJT, UWE	Auspos, Cave, and Long (1988)	XL
NJGD	DEM	1984-1987	AFDC recipients	OJT	Freedman, Bryant, and Cave (1988)	XL
MFSP	DEM	1982-1987	Low -income minority single mothers	CT, OJT	Burghardt et al. (1992) Zambrowski et al. (1983)	XL
ET	DEM[b]	1983-1989	AFDC recipients	CT, UWE[d]	Nightingale et al. (1991)	NXL
JS	DEM	1985-1988	High school dropouts	CT	Cave et al. (1993)	XL
NC	DEM	1989-1992	AFDC high school dropouts	CT, PWE, UWE	Quint et al. (1994)	XL
JTPA	NAT	1983-present	Disadvantaged adults and youth	CT, OJT	Orr et al. (1996)	XL

[a]JOBS68 was a national program but is classified as a demonstration because it was short-lived and featured only a single training activity.

[b]ET was a state-run version of a national program but is classified as a demonstration because its research interest lies mainly in the large scale of its voluntary approach to training for welfare recipients.

[c]Most programs with training components also provided assistance with job search.

[d]Other services were also provided including job development and college assistance.

Key: MDTA= Manpower Development and Training Act; NYC= Neighborhood Youth Corps; JOBS68= Job Opportunities in the Business Sector; JC= Job Corps Program; S-D = Seattle-Denver Income Maintenance Experiment; CETA= Comprehensive Employment and Training Act; SW= National Supported Work Demonstration; HHA= AFDC Homemaker-Home Health Aide Demonstrations; TOPS= Maine Training Opportunities in the Private Sector Program; NJGD= New Jersey Grant Diversion Project; MFSP= Minority Female Single Parent Demonstration; ET= Massachusetts Employment and Training Choices Program; JS= JOBSTART Demonstration; NC= New Chance Demonstration; JTPA= Job Training Partnership Act; NAT= national; DEM= special demonstration; AFDC= Aid to Families with Dependent Children; CT= classroom training (basic education and occupational skills training); OJT= on-the-job training; UWE= unpaid work experience; PWE= paid work experience; NXL= nonexperimental; XL= experimental.

As previously indicated, the measure of training program effects – the dependent variable in the meta-analysis – is the effect on earnings.[8] Thus, the estimated effects on earnings were extracted from each of the evaluation reports listed in Table 1 for each year for which they were available. In most studies, annual effects were available, but when they were not (for example, quarterly or semi-annual effects were reported), we converted them to annual equivalents. The number of post-training years over which the earnings effects were estimated varies substantially among the evaluations, but is typically between one and three years and is as much as five years for a few evaluations. In addition, the level of statistical significance of each estimated effect and (when it was available) the sample size on which each estimate was based were also extracted from the evaluation reports.

Training programs may, of course, potentially affect numerous outcomes in addition to earnings, including employment status, welfare and unemployment compensation payments, crime rates, feelings of satisfaction, and so-forth. We focus on earnings effects because a major objective of all government-sponsored training programs is to increase the earnings of participants. Consequently, earnings effects are estimated in many more evaluations of training programs than are other program effects. Moreover, estimates of earnings effects are measured in the same metric, dollars, and, hence, are readily pooled across evaluations, although there are issues concerning appropriate adjustments for inflation over time and cost of living differences across study sites.

As discussed above, the variance of each estimate training program earnings effects is required for weighting the individual studies. Many evaluations of training programs report exact measures of the

---

[8]One of the studies listed in Table 1, Orr et al. (1996), computed two alternative sets of earnings effects estimates for a single program (JTPA). Both sets provided separate estimates for four groups: adult men, adult women, male youths, and female youths. The first set provided separate estimates for each of the sites included in the study – 16 for adult men and women and 15 for male and female youths. The second set was based on samples that were pooled across the individual study sites. However, unlike the site-specific results, separate estimates were obtained for subgroups of individuals, including three training categories (classroom training, on-the-job training, and other types of training), whether or not the trainee was a welfare recipients (adult females only) and whether or not the trainee had been arrested prior to random assignment (male youths only). In our preliminary statistical work, we obtained similar findings regardless of whether the site-specific or the subgroup set of estimates were used. Thus, to minimize the number of results we present in this report, all the findings are based on only the subgroup set of estimates.

statistical significance of the earnings effect estimates (for example, standard errors, t-values, or p-values) that can be readily converted into the required variance measure. Unfortunately, however, some of the evaluations do not report this information. Instead, they merely indicate whether the earnings effect estimate exceeds the 1-percent, 5 percent, or 10 percent levels of statistical significance. In these instances, it was necessary to impute the variance of the earnings effect estimate on the basis of the reported statistical significance level.

These imputations were based on assuming that the p-value of the estimated earnings effect was located at the midpoint of the possible range. Thus, for example, it was assumed that if the level of statistical significance was reported to exceed the 5 percent level but not the 1-percent level, the p-value equaled .03 (the midpoint between .01 and .05). Similarly, if the level of statistical significance exceeded 1-percent, it was assumed that p = .005 (the midpoint between zero and .01); if the level of statistical significance exceeded 10 percent but not the 5 percent level, it was assumed that p = .075 (the midpoint between .05 and .1); and finally if the earnings effect estimate was not statistically significant, it was assumed that p = .3 (the midpoint between .1 and .5).

c. The Explanatory Variables

A number of variables are used to try to explain variation in program earnings effect estimates. In describing these variables, we suggest possible hypotheses about how some of them might have influenced the effect of voluntary government funded training programs on earnings. With the exception of information on program cost, which could not be obtained for a few of the programs, data were available to construct the explanatory variables for all of the programs listed in Table 1. The data required to construct all the explanatory variables other than those pertaining to area economic conditions were extracted directly from the evaluation studies listed in Table 1. As described below, the data needed to construct measures of area economic conditions were obtained from reports published by the federal government.

Program Characteristics.

Each of the programs listed in Table 1 offered a unique set of services to those who were enrolled. Moreover, different enrollees in a given program received different amounts of each of the provided services. Indeed, individual enrollees often received only a subset of the provided services. In principle, if we had the necessary data on individual program enrollees, we could measure the amount of each service that each received. However, we do not have data at the individual enrollee level, but only at the program level or, in the case of some programs, for subgroups of enrollees who received different combinations of services. Thus, the measures of program characteristics that we were able to construct are far from ideal. Nonetheless, they do capture some of the basic differences among training programs.

To develop our key measure of program characteristics, we first made a basic distinction between classroom training and training that occurs at the workplace. We then defined three categories of classroom training, consisting of basic education, classroom skills training, and a combination of basic education and skills training ("CT + basic ed") and two categories of workplace training, consisting of on-the-job training (OJT) and subsidized work (which is sometimes called "paid work experience"). A sixth category was added for programs that provide both classroom training and workplace training. For purposes of the meta-analysis, each of the earnings effect estimates was assigned to one of the six training types. A set of six dummy variables, each representing one of the training categories, is used in the regression analysis to examine how different types of training influence the effect of government-funded training programs on earnings (not all groups utilized all six types). We had no prior expectations about which types of training would be relatively more effective in increasing earnings and which would be relatively less effective.

Different training type categories may really represent different training programs. That is, programs that offer a particular type of training may be more or less successful in increasing the earnings of participants than programs that offer other types of training, but for reasons other than the type of training they provide (for example, they may be well or poorly administered). To control for this possibility, we constructed a set of dummy variables that indicates the training program to which each earnings effect estimates pertains. This set of variables is used in combination with the type of training variables in two of the regressions that we estimated.

Some types of training are likely to be more expensive than other types of training. Consequently, it is possible that certain types of training are more successful in increasing earnings than other types because more resources are spent in providing them. To control for this possibility in examining the effects of different types of training, and also to learn something about the relation between program cost and program effectiveness, we constructed a variable that measures the program cost per participant.[9] We anticipated that, after controlling for other factors, such as type of training, more costly programs would be more successful in increasing the earnings of participants than less costly programs. Unfortunately, program cost per participant is not available for adults for the Job Opportunities in the Business Sector (JOBS68) program and for youth for both the CETA and Neighborhood Youth Corps (NYC) programs.[10] To deal with this problem in the regression analysis, we set program cost to zero for these programs and included a dummy variable equaling one for these programs and zero for the remaining programs.

Program Enrollee Characteristics.

Unfortunately, the evaluation studies listed in Table 1 do not provide a consistent set of measures of the characteristics of program enrollees (e.g., average age, average educational achievement, and marital status). However, many of the studies (but not all) estimated separate program earnings effects

---

[9]The quality and completeness of cost information varies considerably across the programs listed in Table 1. To construct the cost measure, we attempted, whenever possible, to obtain net administrative cost per participant (excluding opportunity costs to trainees), subtracting out training costs expended on comparison group members. Such estimates represent the additional real resources consumed by the program. Net cost estimates, however, are often not available in the studies that used a non-experimental methodology. Moreover, for some kinds of training (primarily paid work experience), we often could not remove payments to participants from the published cost estimates.

[10]Some of the evaluation reports for the remaining programs that are listed in Table 1 also do not provide estimates of program costs. However, these programs were all subjected to two or more evaluations, at least one of which does provide cost information. As a result, we have at least one cost estimate for each of them.

for white and minority group enrollees. In addition, the studies of youth enrollees usually estimated separate earnings effects for boys and girls. Thus, a set of three dummy variables was constructed indicating whether each earnings effect estimate pertains to white trainees, non-white trainees, or a mixed group of both white and non-white trainees.[11] In addition, a second set of three dummy variables was constructed indicating whether each of the earnings effect estimates for youths pertains to boys, girls, or a trainee group that includes both boys and girls. Although we did not have prior expectations as to whether training programs for youths are more effective for boys or for girls, two opposing hypothesis about the relation relationship between race and training program effects on earnings can be formulated. On the one hand, because white training program enrollees tend to have higher levels of formal education, to live in neighborhoods that are more accessible to jobs, and to be less subject to discrimination than program enrollees from non-white groups, program earnings effects could be larger for white enrollees than for racially mixed groups of enrollees but smaller for minority enrollees. On the other hand, white workers may be better able to succeed in the labor market on their own without the aid of training than non-white workers. If so, program effects on earnings will be larger for non-whites than for racially mixed groups and smaller for whites.

Area Economic Conditions.

The effects of government funded training programs on earnings are likely to be influenced by the economic conditions that prevailed at the times and places the training took place. The evaluations that we use in the meta-analysis measured the effects of training programs on earnings under a wide variety of economic conditions. Although each individual study estimated training effects over only a few years, taken together, they cover a time span of more than three decades. Moreover, although some of

[11] In most instances, earnings effect estimates for black, Hispanic, or Asian program participants, or a mix of persons from these groups, are treated as pertaining to trainees from a "minority" group. However, a few evaluation studies distinguish between black and non-black enrollees, rather than white and minority enrollees. In these cases, earnings effect estimates for the non-black group, which may include some Hispanics and Asians, were treated as pertaining to "white" trainees.

the evaluations examined the effects of training on a nationally representative sample of trainees, most were limited to measuring training effects in one or more specific local communities, and these communities differed across studies.

We use three variables to investigate the influence of economic conditions on training program earnings effects at the time and at the place each effect of training on earnings was estimated: the unemployment rate, the percent of the workforce in manufacturing, and (for women only) the maximum AFDC payment for a family of three.

In theory, the unemployment rate could be either positively or negatively related to the effect of training programs on earnings. On the one hand, if the unemployment rate is low, then those who receive training may readily find jobs, but so may similar persons who did not receive training; however, those who receive training may enjoy a competitive advantage over similar persons who did not when the unemployment rate is high and jobs are difficult to find. If so, the relationship between the unemployment rate and the earnings effect would be positive. On the other hand, training may only be helpful if unemployment is low and the program can help prepare individuals for the relatively large number of available jobs; if unemployment is high, it may be very difficult for trainees, as well as other workers, to obtain a job, and the effects of training on earnings may be negligible. Consequently, the relationship would be negative.

We anticipated that the percentage of the workforce in manufacturing is positively related to the effects of training programs on earnings. Traditionally, manufacturing jobs pay low-skilled workers higher wages and are otherwise more attractive than jobs in the service industry. Thus, a proportionately high number of jobs in manufacturing should serve both to motivate low-skilled persons to seek training and employment and to reward them with higher earnings once they find a job.

We obtained the data on unemployment rates and the percentage of the workforce in manufacturing for the dates and locations we required from a variety of U.S. government publications such as the *Monthly Labor Review*, *Employment and Earnings*, the *Economic Report of the President*, and

County Data Patterns and from the U.S. Bureau of Labor Statistics' web page at *http://www.bls.gov/*. Many of the estimates of program earnings effects that we use in the meta-analysis pertain to trainees in two or more local communities. In such instances, values for the unemployment rate and the percentage of the workforce in manufacturing were obtained for each site and a weighted average was computed, with the percentage of the evaluation sample contributed by each site used as weights.

Because each of the program earnings effect estimates are for a particular age and gender group and often for a specific racial group, the unemployment rate variable that we constructed is also age-, gender-, and race-specific. However, although national unemployment rates are available by age, gender, and race, local unemployment rates are usually only reported for the entire local labor force. Thus, to obtain age-, gender-, and race-specific local unemployment rates, each reported local unemployment rate was multiplied by the ratio of the national unemployment rate for the group of interest to the overall national unemployment rate.

AFDC payments might influence the size of the earnings effect for women. For purposes of this study, we use the maximum AFDC payment for a family of three in the state in which the training took place. Information on this variable was obtained from various issues of *The Green Book*, which is produced annually by the staff of the Committee on Ways and Means of the U.S. House of Representatives and available on the web at *http://aspe.hhs.gov/20000gb/index.htm*.

In principle, the maximum AFDC payment for which a family of three is eligible may be either positively or negatively related to the effect of training on the earnings of adult women. On the one hand, AFDC is a potential alternative to work for single parents. Hence, generous AFDC payments may reduce the incentive of some adult women to seek training and jobs and, consequently, could reduce the effectiveness of training programs. On the other hand, more generous AFDC payments tend to be found in wealthier, higher wage states; and training may have larger effects on earnings in such states.

Evaluation Method.

As indicated in Table 1, some of the earnings effect estimates that are used in the meta-analysis are based on a random assignment methodology, while others were obtained by using non-experimental methods. Non-experimental methods produced a number of estimates that were much more negative than any of the estimates produced by the experimental methods and numerous others that were much more positive, suggesting that the use of non-experimental methods can result in substantial estimation errors. To control for this possibility in the regressions analysis, we use a dummy variable that is set equal to one for studies that used an experimental evaluation design and zero for studies that used a non-experimental design. We did not have prior expectations about whether the relation between this variable and training program effects on earnings is positive or negative.

Years Since Training.

Most of the studies listed in Table 1 measured program effects on earning for the first, second, and/or third year after the training was received. However, a few studies measured these effects during the fourth and even the fifth post-training year. Thus, a variable was constructed equaling one for program earnings effect estimates that pertain to the first post-training year, two for earnings effect estimates that pertain to the second post-training year, and so forth.

This "years since training" variable is used to test two polar opposite hypotheses about how the effects of government funded training programs change over time. The first hypothesis is based on the frequent suggestion that although such programs may initially give workers who participate a competitive advantage in the labor market, this advantage will diminish over time as similar workers who did not receive the training catch up. The second and contrary hypothesis is based on the possibility that training may open doors that allow participants to obtain additional training after they leave the program and take a job. Consequently, the program's effects on earnings will grow over time. If the first of these hypotheses is valid, the years since training variable will be negatively related to the earnings effect estimates, but if the second hypothesis is correct, it will be positively related.

Calendar Year of Training.

The earliest year in which any of the program enrollees included in the studies listed in Table 1 received training was 1964. We constructed a simple time trend variable that equals one for training received in 1964, two for training received in 1965, three for training received in 1966, and so forth. The purpose of this variable is to test whether government funded training programs have improved over time because more has been learned about running them. If so, the relation between the year of training variable and the earnings effect estimates should be positive.

4. Descriptive Statistics of the Sample

As we have indicated, our sample contains estimated training effects from 31 studies of 15 voluntary programs that ran in the United States between 1964 and 1998. Because two of the studies (Kiefer, 1978 and Gay and Borus, 1980) each evaluate 4 programs, we have 37 separate evaluations of 15 programs.

Many of the studies produced estimates for different population groups (men, women, youth, black, white, welfare, non-welfare, etc.) and different time periods after training (between one and five post-training years). In total, there are 315 estimates, 83 for men, 133 for women, and 99 for youth. The number of estimates for each study is reported in Table 2.

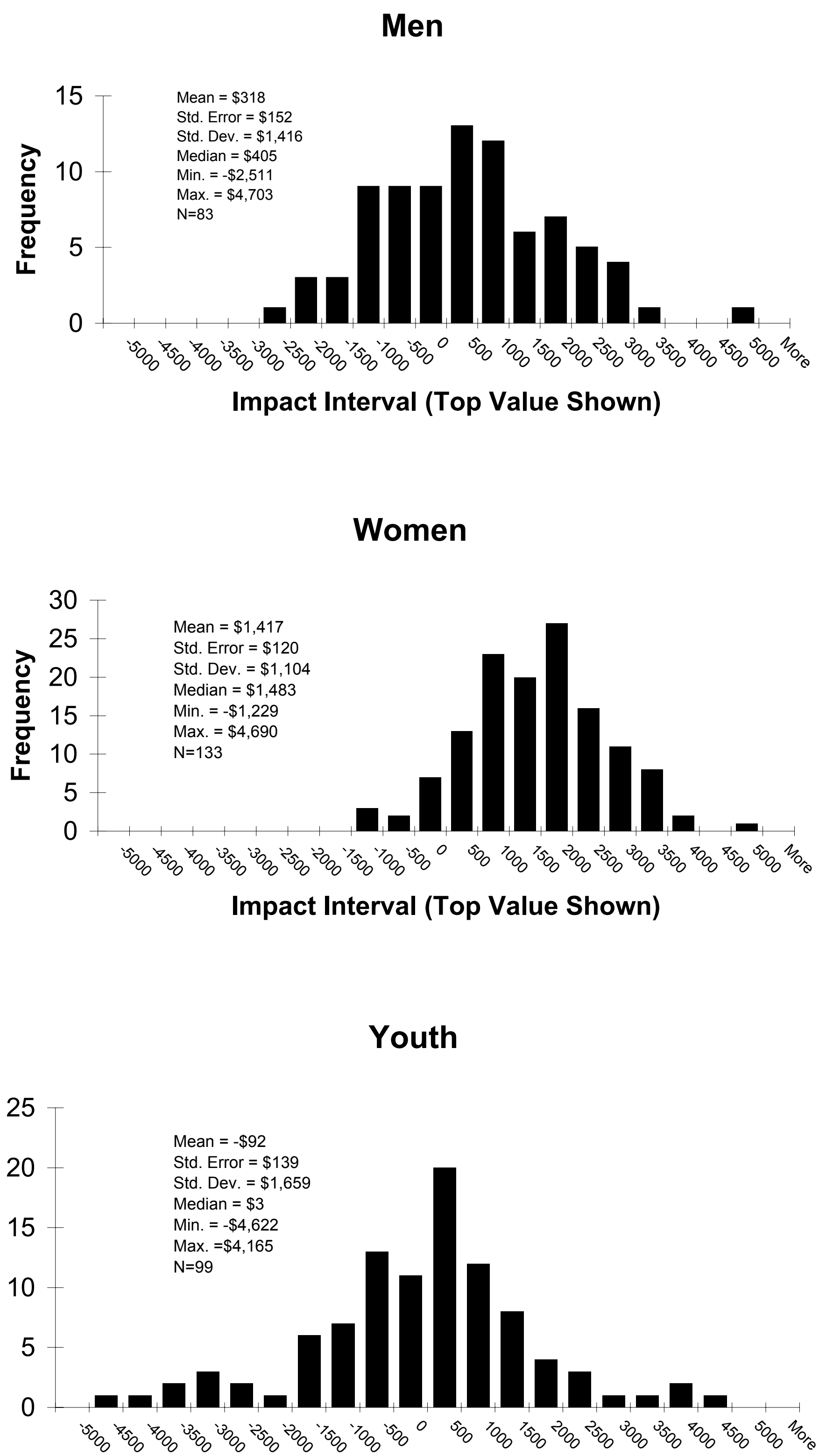
All but 4 of the studies listed in Table 2 produced more than one estimate. Kiefer (1979) produced the most estimates, with 48 (15 percent of the sample), while the average the number of estimates per study is just over 10 (315/31). Seventeen studies produce estimates for men, averaging just under 5 estimates per study, 23 studies produce estimates for women, averaging just under 6 estimates per study (133/23), and 15 studies produce estimates for youth, averaging just over 6.5 estimates per study (99/15)

Figure 1 presents histograms of the estimated earnings effects for the three major population groups (men, women, and youth). Figure 1 also shows the mean (unweighted), the standard error of the estimated mean (unweighted), the standard deviation (unweighted), the median, the minimum, and the maximum.

Table 2
Number of Estimates by Study

Study	Men	Women	Youth	Total
Ashenfelter (1978)	10	10	0	20
Ashenfelter and Card (1985)	1	1	0	2
Auspos et al. (1988)	0	2	0	2
Bassi (1983)	6	12	0	18
Bassi (1984)	4	4	6	14
Bell and Orr (1994)	0	14	0	14
Bloom (1984)	10	10	0	20
Bloom (1987)	3	3	0	6
Bryant and Rupp (1987)	6	6	3	15
Burghardt et al. (1992)	0	8	0	8
Cain (1968)	0	0	1	1
Cave et al. (1993)	0	0	9	9
Cooley (1979)	3	3	0	6
Couch (1992)	0	5	5	10
Dickinson and West (1983)	2	4	0	6
Dickinson et al. (1984)	0	0	9	9
Dickinson et al. (1986)	3	3	0	6
Dickinson et al. (1987a)	3	3	0	6
Dickinson et al. (1987b)	0	0	2	2
Freedman et al. (1988)	0	2	0	2
Gay and Borus (1980)	4	4	8	16
Hollister et al. (1984)	4	1	2	7
Kiefer (1978)	2	0	0	2
Kiefer (1979)	12	12	24	48
Mallar et al. (1982)	0	0	12	12
Nightingale et al. (1991)	0	7	0	7
Orr et al. (1996)	6	12	12	30
Quint et al. (1994)	0	0	1	1
Schochet et al. (2000)	0	0	2	2
Westat (1984)	4	6	3	13
Zambrowski and Gordon (1993)	0	1	0	1
Total	83	133	99	315

Figure 1
Histograms of Training Effects by Group



As the histograms indicate, there is considerable variation in the estimated training effects. Some of this variation is undoubtedly due to differences in the programs, but some may also be due to differences among the population groups studied, the time periods studied, economic conditions in the places in which the studied programs were located, and the methodologies used to evaluate specific programs.[12] In addition, as will be discussed in detail below, some of the variation is due to sampling error and some is due to random error arising from unmeasurable factors.

Among the three groups, women have by far the highest mean training effect of $1,417, which is statistically significant at the 1-percent level. Men have a mean effect of $318, which is statistically significant at the 5 percent level, and youth have a negative mean effect of -$92, which is not statistically different from zero. For all groups, the median is fairly close to the mean, indicating that the distributions are fairly symmetrical. Thus, the implication is that training works best for women, works modestly well for men, and is ineffective for youth.

In addition to being most effective for women, the distribution of training effects is also narrowest for women, with a standard deviation of $1,104 and a range from -$1,229 to $4,690. The distribution is most spread out for youth, with a standard deviation of $1,659 and a range from -$4,622 to $4,165. Thus, while training appears ineffective for youth, there is also great uncertainty associated with this conclusion.

Table 3 presents the mean earnings effect for each training program, along with a percentage breakdown of the sample by training program. For men, the estimated effects vary from a low of -$1,805 for SIME/DIME (statistically significant at the 10 percent level) to a high of $1,310 for Supported Work (statistically significant at the 10 percent level). Almost 80 percent of the estimates are for the MDTA and CETA programs, with MDTA having an average effect of $642 (statistically significant at the 1-

[12]For example, six of the studies evaluated the CETA program. Because all of the CETA evaluations were non-experimental, the methodologies used by the authors differed across the studies and the estimates for a particular population group vary considerably. For a discussion of the CETA evaluations, see Barnow (1987).

Table 3
Unweighted Mean Effects of Training by Program and Group

	Men		Women		Youth	
	Mean	**Fraction of Sample**	**Mean**	**Fraction of Sample**	**Mean**	**Fraction of Sample**
Overall Mean	318 **	1.00	1417 ***	1.00	-92	1.00
	(152)		(120)		(139)	
Mean by program						
MDTA	642 ***	0.40	2159 ***	0.23	--	--
	(236)		(176)			
SIME/DIME	-1805 *	0.02	525	0.03	--	--
	(960)		(489)			
SupportedWork	1310 *	0.05	860 **	0.05	28	0.07
	(679)		(399)		(607)	
CETA	6	0.36	1346 ***	0.29	698 **	0.23
	(248)		(159)		(335)	
JOBS68	-149	0.10	2069 ***	0.06	--	--
	(480)		(346)			
JTPA	761	0.07	1179 ***	0.09	-357	0.12
	(554)		(282)		(464)	
Massachusetts ET	--	--	501	0.05	--	--
			(370)			
Home Health Aide	--	--	1558 ***	0.11	--	--
			(261)			
NJ OJT	--	--	1088	0.02	--	--
			(692)			
Maine TOPS	--	--	1077	0.02	--	--
			(692)			
MFSP	--	--	309	0.07	--	--
			(326)			
Job Corps	--	--	--	--	-221	0.31
					(289)	
NYC/OS	--	--	--	--	-1055 ***	0.16
					(402)	
Jobstart	--	--	--	--	332	0.09
					(536)	
New Chance	--	--	--	--	-308	0.01
					(1607)	
F statistic for differences in programs	2.41		4.60		2.09	
p-level for F-test	0.044		0.000		0.062	
Total Sample Size	83		133		99	

Note: Standard erors in parentheses.
* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

percent level) and CETA an average effect of $6 (not statistically significant). For women, the estimates vary from a low of $309 for MFSP (not statistically significant) to a high of $2,159 for MDTA (statistically significant at the 1-percent level). About half of the estimates for women are for the MDTA and CETA programs and in both cases the effects are fairly large and statistically significant at the 1-percent level. For youth, the estimates vary from a low of -$1,055 for NYC/OS (statistically significant at the 1-percent level) to a high of $698 for CETA (statistically significant at the 5 percent level). About a third of the estimates for youth are for the CETA program and about a fourth are for the Job Corps program.

The F-test that is reported at the bottom of Table 3 indicates that, for all three groups, we can reject the hypothesis that the effects are the same for each program. The test is only marginally rejected for youth (p-level=.062), moderately rejected for men (p-level=.044), and strongly rejected for women (p level=.000). Thus, based on these tests, we conclude that some of the variation in the effects is due to differences in the programs. In what follows, we specify a statistical model that attempts to isolate some of the factors responsible for the differences across programs.

5. A Model Explaining Variation in Training Effects

In meta-analysis, two types of statistical models have been used to isolate the effects of various factors that cause program effects to vary across studies. These models have been termed "fixed effect" and "random effects" models, although as will be discussed below, the latter is really a generalization of the former and is more appropriately termed a "mixed effect" model.

Both the fixed and random effects models take into account the fact that the individual underlying earnings effect estimates are based on different sample sizes, and hence have different levels of statistical precision. As previously suggested, it would not make sense to weigh two studies equally that produce estimates having very different levels of statistical precision. For example, suppose one study produced an estimated effect of a particular training program of $5,000, but this estimate was very imprecise and not statistically significant because of a small sample size of, say, only 500 persons. Suppose another study produced an estimated effect of $1,000 for the same program, but was very precisely estimated because of a much larger sample of, say, 4,500 persons. If we did not take into account the sample sizes, we would conclude that the average effect of the program was $3,000 (the unweighted mean of the estimates produced by the two studies). However, the average effect is probably closer to $1,000 because of the total sample used in the two studies (5,000), 90 percent was from the latter study.

In order to account for sampling variation across studies, the following statistical model is specified:[13]

(1) $T_i = T_i^* + e_i$, where

T_i is the estimated training program effect, T_i^* is the "true" training program effect (obtained if the entire target population was evaluated), and e_i is the error due to estimation on a sample smaller than the population. It is assumed that e_i has a mean of zero and a variance of v_i.

In order to provide an estimate of the mean effect that takes into account the fact that v_i varies across studies (that is, v_i is smaller for studies with larger samples), a weighted mean can be calculated, with the weight being the inverse of the v_i, $1/v_i$. If sampling variation is the only source of variation in the training program effects, weighting in this manner produces the most precise estimate of the mean program effect.

Using the estimated variances from each study produces a weighted mean estimate of $471 for men (compared to an unweighted mean of $318), a weighted mean estimate of $832 for women (compared to an unweighted mean of $1,417), and a weighted mean of -$28 for youth (compared to an unweighted mean of -$92). Thus, except for women, the weighted means are close to the unweighted means. The weighted mean is statistically different from zero at the 1-percent level for men and women and is not statistically significant for youth.

---

[13]The remainder of this section is drawn from Raudenbush (1994).

Sampling variation is not the only source of variation in estimates across studies, however. There are two other sources of variation that are taken into account in meta-analysis studies. One source has to do with the fact that the estimates are produced for different programs, over different time periods, for different population groups, in different locations, and so forth. The other results because there are unmeasurable factors that cause variation in program effects. These could be related to staff attitudes toward training participants and other features of the program or environment that are inherently unmeasurable.[14]

Each of these sources of variation may be identified by extending the model described by equation (1) in the following way:

(2) $T_i^* = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots\beta_p X_p + u_i,$

where β_0 is the model intercept, the Xs are observed characteristics of the studies that cause variation in the true program effects T_i^* , the βs are coefficients representing the marginal effects of the characteristics on the true program effect, and u_i is a random error term with variance σ^2, representing unmeasured factors causing variation in program effects. Equation (2) is sometimes termed a "structural" model in the meta-analysis literature.

Together, equations (1) and (2) constitute a statistical model of the variation in program effects. Substituting equation (2) into equation (1) yields the mixed effect model we estimate:

(3) $T_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots\beta_p X_p + e_i + u_i.$

In equation (3), there are three potential sources of variation in T_i – sampling error (the e_i), observed characteristics of the studies (the Xs), and random error (the u_i). If the βs are not zero, but u_i is identically zero for all studies, then the model is referred to as a "fixed effects" model. In the fixed effect

[14]Theoretically, it might be possible to collect information on many of the factors seem unmeasurable. However, most studies of training programs do not produce a consistent set of information on program practices and other difficult to measure factors. The most ambitious attempt to collect such factors is described in the study by Bloom, Hill, and Riccio (2001). However, this study is unique in that all of the underlying evaluations were performed by the same organization (MDRC) and utilized a common set of survey instruments.

model, there are two sources of variation in the estimated program effects – sampling error and variation in observed characteristics. The weight used in estimating the fixed effects model is the inverse of the sampling variance ($1/v_i$), because the only source of variation in the estimates, other than the Xs, is the sampling variance. If the βs are not zero and u_i, as well as e_i, varies across studies, then the model is referred to as a "mixed effects" model. In the mixed effects model, there are three sources of variation in the estimated program effects – sampling error, variation in observed characteristics, and random error caused by variation in unobserved characteristics. The weight used in estimating the mixed effects model is the inverse of the sum of the sampling error plus the random effects error ($1/[v_i + \sigma^2]$). Clearly, the fixed effects model is a special case of the mixed effects model. It is possible to test statistically for the significance of the fixed and random effects.

To estimate the mixed effects model, an estimate of σ^2 is obviously needed. Raudenbush (1994) describes a variety of procedures for estimating the model, including method of moments estimators and maximum likelihood estimators. One procedure, based on a method of moments estimator, involves the following steps. First, equation (2) is estimated using ordinary least squares (OLS). Then, the mean square residual variance from the regression is used to calculate an estimate of σ^2, based on the following formula:

(4) $\hat{\sigma}^2 = \text{MSR} - k/(n-p-1)$,

where MSR is the mean square residual from the OLS regression[15] and k is a constant given by the following formula (see Raudenbush, 1994, p. 319):

(5) $k = \Sigma v_i - \text{trace}[\mathbf{X'VX(X'X)}^{-1}]$,

where the boldface refers to matrix notation for the vector of p explanatory variables (the X_i) and the n sampling variances (the v_i), and trace is the sum of the diagonal elements of the resulting matrix. Essentially, the estimate of σ^2 is based on the total residual variance from the OLS regression less an

[15]The MSR is calculated by dividing the residual sum of squares by the number of degrees of freedom in the regression, which is the number of observations (n) minus the number of βs estimated in the model (p+1).

adjustment term based on a weighted average of the sampling errors (v_i) for each observation. After obtaining the estimate of σ^2, the model is re-estimated by weighted least squares, using $1/[\hat{\sigma}^2 + v_i]$ as weights.

Using the estimated variances (v_i) from each study and the method of moments estimator of σ^2 described by equations (4) and (5) as weights produces a mixed effect estimate of $337 for men (compared to a fixed effect mean of $471), a mixed effect estimate of $1,422 for women (compared to a fixed effect mean of $832), and a mixed effect mean of -$27 for youth (compared to a fixed effect mean of -$28). Thus, for men and women, the mixed effect means differ from the fixed effect means (but are similar to the unweighted means), while for youth, the mixed effect mean is virtually identical to the fixed effect mean. The mixed effect mean is statistically significant at the 1-percent level for men and women, but is not statistically significant for youth.

In addition to the fixed and mixed effects models, there is a third model, called the "unweighted model," in which it is assumed that there is no variation in the v_i across studies. If all studies have the same sample sizes, then the unweighted model is appropriate and can be estimated by a simple ordinary least squares regression of the program effects on the observed characteristics. Of course, if standard errors of the program effects are not available for the studies, the unweighted model must be used.[16] Sometimes, however, the unweighted model is appropriate if there is uncertainty about the accuracy of the estimated standard errors from the underlying studies. For example, as discussed earlier, we had to impute standard errors for some of the studies in our sample because the authors provided only broad significance levels (10 percent, 5 percent, 1-percent, or not significant).[17] Furthermore, and perhaps most

---

[16]An alternative to using the standard errors to weight the observations is to use the sample sizes (there is a correspondence between the standard error and the sample size). However, sample sizes were frequently missing for the studies we use and, even when the total sample size (including both the trainees and the comparison groups) was provided, sample sizes for the trainee sample were often not available.

[17]We had to impute standard errors for 71 of the 315 cases (23 percent of the sample). The standard error was missing 17 times for men (20 percent of the sample), 22 times for women (17 percent of the sample), and 32 times for youth (32 percent of the sample).

importantly, when it is only known that the estimate is not statistically significant, there is considerable uncertainty in estimating the standard error because the true significance level lies between 10 percent and 100 percent.[18]

For completeness, we estimated all three models (the unweighted, fixed effects, and mixed models). Using the test suggested by Raudenbush (1992, p. 314), the unweighted and fixed effect models were emphatically rejected in favor of the mixed model for almost every model specification for all three groups.[19] Accordingly, in the tables that follow, we only present results from the mixed model. However, the reader should be aware that the test favoring the mixed model is based on the assumption that the standard errors derived from the studies are accurate. In the appendix, we present the full set of estimates for all three models.

6. Results

The model given by equation (3) is estimated separately for men, women, and youth, using the set of covariates defined earlier. Table 4 presents separately for men, women, and youths the means and standard deviations of these covariates in our sample of studies. Most of the earnings estimates are for either non-white trainees or a mixed group of white and non-white trainees (70 percent for men, 76 percent for women, and 79 percent for youth). In addition, most of the estimates are non-experimental (86 percent for men, 63 percent for women, and 69 percent for youth). The average unweighted unemployment rate in the study sites was 5.5% for men, 6.9% for women, and 22.3% for youth. A bit under one-quarter of the workforce in the labor markets in which the programs operated were employed in manufacturing. The average earnings effect is for training that occurred about two years earlier. The

[18]The significance level was above 10 percent in 43 of the 71 cases in which a standard error was imputed (14 percent of the total sample).

[19]The test for the mixed effect model is a test of the hypothesis that sigma-square is zero. The test statistic is given by $Q=\Sigma w_i(T_i - \beta_0 - \beta_1 X_1 - \beta_2 X_2 - .....\beta_p X_p)^2$, where $w_i = 1/v_i$. This statistic is approximately distributed as chi-square, with n-p-1 degrees of freedom.

Table 4
Unweighted Means and Standard Deviations of Covariates

	Men (N=83)	Women (N=133)	Youth (N=99)
Training Type			
Classroom skills training	0.51	0.38	0.07
	(0.50)	(0.49)	(0.26)
Basic education	0.00	0.02	0.00
	(0.00)	(0.15)	(0.00)
CT+Basic ed	0.00	0.03	0.01
	(0.00)	(0.17)	(0.10)
OJT	0.18	0.14	0.07
	(0.39)	(0.35)	(0.26)
Subsidized work	0.11	0.13	0.16
	(0.31)	(0.34)	(0.37)
Mix of classrooom and workplace	0.20	0.29	0.69
training	(0.41)	(0.46)	(0.47)
1=female youth	0.00	0.00	0.47
	(0.00)	(0.00)	(0.50)
1=males and female youth combined	0.00	0.00	0.08
	(0.00)	(0.00)	(0.27)
1=Experimental dummy	0.14	0.37	0.31
	(0.35)	(0.48)	(0.47)
1=White[a]	0.30	0.24	0.21
	(0.46)	(0.43)	(0.41)
1=Nonwhite[a]	0.40	0.31	0.23
	(0.49)	(0.46)	(0.42)
Unemployment rate (%)	5.55	6.95	22.27
	(2.47)	(2.47)	(9.28)
Unemployment rate squared	36.80	54.35	581.29
	(31.80)	(37.62)	(518.51)
Percent manufacturing employment	24.77	23.07	21.04
	(3.31)	(4.97)	(4.10)
Years since training	2.17	1.99	2.01
	(1.05)	(1.00)	(0.93)
Year of training (1964=0)	7.90	11.76	12.37
	(6.60)	(7.64)	(7.46)
Program cost[b]	7,080	6,591	8,782
	(3,573)	(3,690)	(4,031)
1=Program cost missing	0.10	0.06	0.42
	(0.30)	(0.24)	(0.50)
Program[c]			
1=MDTA	0.40	0.23	0.00
	(0.49)	(0.42)	(0.00)
1=SIME/DIME	0.02	0.03	0.00
	(0.15)	(0.17)	(0.00)
1=SupportedWork	0.05	0.05	0.07
	(0.22)	(0.21)	(0.26)
1=CETA	0.36	0.29	0.23
	(0.48)	(0.45)	(0.42)
1=JOBS68	0.10	0.06	0.00
	(0.30)	(0.24)	(0.00)
1=JTPA	0.07	0.09	0.12
	(0.26)	(0.29)	(0.33)
1=Massachusetts ET	0.00	0.05	0.00
	(0.00)	(0.22)	(0.00)
1=Home Health Aide	0.00	0.11	0.00
	(0.00)	(0.31)	(0.00)
1=NJ OJT	0.00	0.02	0.00
	(0.00)	(0.12)	(0.00)
1=Maine TOPS	0.00	0.02	0.00
	(0.00)	(0.12)	(0.00)
1=MFSP	0.00	0.07	0.00
	(0.00)	(0.25)	(0.00)
1=Job Corps	0.00	0.00	0.31
	(0.00)	(0.00)	(0.47)
1=NYC/OS	0.00	0.00	0.16
	(0.00)	(0.00)	(0.37)
1=Jobstart	0.00	0.00	0.09
	(0.00)	(0.00)	(0.29)
1=New Chance	0.00	0.00	0.01
	(0.00)	(0.00)	(0.10)

[a]Omitted category is mixed race/ethnicity.
[b]Average program cost is given over non-missing values.
[c]Totals within groups may not add to one because of rounding.

unweighted average program cost (in 1999 dollars) was \$7,080 for men, \$6,591 for women, and \$8,782 for youth. Thus, the largest earnings effects were for the group having the lowest average cost (women) and the smallest earnings effects were for the group having the largest average cost (youth).

Because of the large number of covariates in the model, we introduce them sequentially in the following order: (1) training type, (2) whether the evaluation was a randomized experiment, (3) race/ethnicity, (4) the unemployment rate and its square, (5) the percent of the local labor force that is in the manufacturing sector, (6) years since the training occurred, (7) the calendar year of the training, and finally (8) program cost.[20] All covariates except the training types and years since training are centered about their means in the sample to help make the results readily interpretable. The training types are simply dummy variables for each training type. Years since training is centered around 1.

Tables 5 - 7 present the results for men, women, and youth, respectively. The tables report the marginal effects for each of the 8 models estimated, their standard errors, and their level of statistical significance. Also presented is the proportion of the total estimated variance of the error term that is due to the random error component (σ^2/v^*, where v^* is the total estimated error variance $\sigma^2 + v$)[21], the percentage of the total variance from the regression that is explained by the covariates ($1 - [v^*/s^2]$, where s^2 is the total variance from the regression), and the probability level (p-level) for the statistical significance of the covariates.[22]

Because all the covariates used to obtain the estimates in Tables 5 - 7 except training type and years since training are centered around the sample means, the coefficients on the training types can be

[20]An additional specification was estimated that included dummy variables for each of the programs. However, due to high collinearity with the other covariates, the coefficients were very imprecisely estimated and the results were largely uninformative. Table A-1 of the appendix presents unweighted, fixed, and mixed effects estimates of a model including only the program variables. Additionally, we examined several models that interacted the training types with other covariates (such as the unemployment rate), but again the results were difficult to interpret because of multicollinearity problems.

[21]The higher this percentage, the closer the estimates will tend to be to the unweighted estimates and the lower this percentage, the closer the estimate will be to the fixed effects estimates.

[22]The p-level is based on an unweighted OLS regression.

interpreted as mean program effects for the average person in each type of training, a year after the training took place.[23] The overall mean training effect for each group is a weighted average of the effects of the individual training types, where the weights are the percentage of observations in each training-type category. As indicated in Table 4, over half of the earnings estimates for men and almost two-fifths of the estimates for women are for classroom skills training, while over two-thirds of the estimates for youth pertain to a mix of classroom and workplace training.

a. Men

Of the three groups we examined, the findings for men tend to be the least robust and the most difficult to interpret, possibly because the number of observations for men is also the smallest. In Models 1 through 3 in Table 5, the most effective type of training for men appears to be classroom skills training. In Model 1, for instance, the effect of classroom skills training is $581. However, as indicated by the high p-level, which appears at the bottom of Table 5, this effect is not significantly different from the effect of the other training types. Indeed, once the unemployment rate is taken into account in Model 4, the effect of classroom skills training becomes small and statistically insignificant. However, the effects of both OJT and mixed classroom and workplace training become positive and statistically significant in Model 4. Subsidized work appears quite ineffective for men. In fact, it exerts a statistically significant negative effect on earnings in Models 4 through 7.

As additional covariates are added to the model in Table 5, the coefficients on the individual training types change substantially. As a result, we cannot draw firm conclusions about the effectiveness of particular types of training for men. However, because none of the coefficients on training types in Table 5 exceeds $1,000, it is apparent that no type of training results in large positive effects on the annual earnings of men. This is, perhaps, not surprising, given the relatively modest overall effect of

[23]The persons who receive one particular type of training may well differ from those who receive some other type of training. Thus, the coefficient on a particular training type indicates whether that type of training is effective for a typical individual who receives it, but a comparison of the coefficients on different training types do not necessarily indicate that one type of training would be superior to another for the *same* individual.

Table 5
Variation in Program Effects, Mixed Effects Model Results
Men

	1	2	3	4	5	6	7	8
Training Type								
Classroom skills training	581 ***	652 ***	636 ***	227	71	26	24	131
	(198)	(205)	(207)	(189)	(157)	(206)	(194)	(404)
OJT	322	308	329	598 **	874 ***	828 ***	848 ***	312
	(359)	(361)	(364)	(301)	(256)	(290)	(274)	(436)
Subsidized work	-171	-387	-277	-193	-771 **	-856 *	-848 **	-961
	(460)	(483)	(512)	(411)	(361)	(440)	(415)	(1267)
Mix of classroom and workplace	57	107	73	530 **	766 ***	717 ***	582 *	465
training	(273)	(277)	(282)	(237)	(187)	(237)	(336)	(413)
Experimental dummy		755	975 *	1475 ***	1568 ***	1583 ***	1342 ***	1218 **
		(479)	(558)	(467)	(394)	(396)	(460)	(616)
White			205	-775 *	-851 **	-835 **	-700 *	-684
			(465)	(426)	(338)	(341)	(400)	(430)
Non-white			405	1217 ***	-84	-222	-162	-326
			(443)	(426)	(453)	(612)	(590)	(630)
Unemployment rate				-1131 ***	-168	-71	44	111
				(320)	(293)	(414)	(402)	(437)
Unemployment rate squared				61 ***	7	2	-6	-8
				(23)	(19)	(24)	(23)	(25)
Percent manufacturing employment					232 ***	247 ***	288 ***	384 ***
					(59)	(73)	(107)	(122)
Years since training						41	58	111
						(123)	(121)	(130)
Year of training							34	91
							(59)	(74)
Program cost								-0.004
								(0.135)
Program cost missing								976
								(647)
Percentage random variance (σ^2/v^*)	51.4%	51.9%	52.2%	33.9%	17.4%	17.5%	13.7%	16.4%
Percentage explained variance ($1-[v^*/s^2]$)	0.5%	0.1%	0.9%	29.3%	43.9%	44.2%	47.1%	46.2%
P-level for significance of covariates	0.598	0.620	0.675	0.000	0.000	0.000	0.000	0.000

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean.
Years since training is centered around 1. P-level given is from unweighted model.
* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

training on men's earnings mentioned earlier. It is also clear from the negative coefficient on subsidized work in Table 5 that this form of training is ineffective. Hence, while it is not so clear which of the other types of training are effective, they are clearly more effective than subsidized work.

A number of interesting findings emerge from examining the coefficients on the remaining covariates reported in Table 5. First, experimental studies tend to produce considerably larger earnings effects for men than non-experimental studies. This finding generally strengthens as additional covariates are added to the model. However, only 14-percent of our 83 observations for males were obtained from experimental studies, and, as will be seen, a similar finding does not occur for either adult women or youths.

Second, in Table 5, the effect of training on white men and on non-white men is implicitly compared to training effects on an omitted category, a mixed group of white and non-white men. Although the non-white coefficient is not very robust to changes in model specification and, hence, difficult to interpret, it seems clear that white men benefit less from training than the mixed and (probably) non-white groups. This effect tends to persist as other covariates are added.

Third, in Model 4, the unemployment rate appears to be nonlinearly related to the size of the training effect, but the relationship largely disappears when the percent of employment in manufacturing is added as a covariate in Model 5, probably because of collinearity between these two variables. The result in Model 4 indicates that the unemployment is negatively related to the training effect at low unemployment rates, and at very high levels of unemployment becomes positively related to the training effect. The effect of the unemployment rate turns positive when the unemployment rate reaches about 14.8 percent, which is well outside the sample range for men (the maximum unemployment rate in the men sample is 10.7 percent). This means that the effect of training declines as unemployment rises, but after reaching a threshold level of unemployment (14.8 percent in this case), the effect begins to rise. However, over the sample range, the effect declines with unemployment. So, for example, in Model 4, the effect of OJT goes from $598 at the mean unemployment rate (5.5 percent) to -$1,848 at an unemployment rate of 8 percent.[24]

Fourth, the effect of training for men appears to be considerably higher in areas where there is more manufacturing employment. Model 8 implies, for example, that a one-percentage point increase in manufacturing employment increases the effect of training by $384. This effect is very robust to model specification. As will be seen, the percent of manufacturing employment does not exert a similar effect on women or youths. Perhaps, this is not surprising, as historically manufacturing has mainly provided high wage jobs for adult men.

Fifth, there is evidence that training effects persist over time after the training has occurred and possibly even increase somewhat. Although the regression coefficient on time since training is not close to being statistically significant in any of the models, its sign is always positive.[25]

Sixth, Models 7 and 8 in Table 5 indicate that there is no statistically significant difference between the effects of more recent training programs, such as JTPA, and the effects of older training programs, such as MDTA Supported Work, and CETA, suggesting that there has been little, if any improvement in the operation of training programs for men over time.

Seventh, for men, there does not appear to be any relationship between the size of the training effect and the cost of the training.

Finally, the estimates for men indicate that there is a substantial random component in explaining variation in the effects of training. In Model 1, for example, just over one-half the estimated error

---

[24]This finding is consistent with the finding of Bloom, Hill, and Riccio (2001). They include only a linear term in the unemployment rate and find that the effect of training declines monotonically as the unemployment rate rises.

[25]An additional specification was estimated with years since training squared in addition to years since training. In this specification, the coefficient of the linear term was 478 and is statistically significant at the 5 percent level (standard error of 231). The coefficient of the squared term was -136 and is also statistically significant at the 5 percent level (standard error of 61). This result implies that the training effects first increase over time (for about the first two years) and then begin to decrease. In further analyses not reported here, we found that the training effects increased over time (from year-to-year) by about the same number of times it decreased. If we focus only on experimental studies, there is strong evidence of increasing training effects over time, while for non-experimental studies there is no pattern in the training effects over time.

variance is due to the random component. However, as might be expected, the random component becomes less important as more covariates are added to the model. In Model 7, for example, less than 20 percent of the error variance is due to the random component. From Model 4 on, the covariates explain a significant portion of the overall variation in the effects of training.

b. Women

The results for women, which are reported in Table 6, imply that there are clear differences across the various training types that are very robust to model specification. Four of the six training types are quite effective. Three of these training types – classroom skills training, OJT, and mixed classroom and workplace training – lead to similar increases in earnings that are well above one thousand dollars per year, while subsidized work results in somewhat smaller, but still substantial, increases. Training involving basic education, on the other hand, appears ineffective. However, only 7 of the 133 observations for women were for this type of training. Over 15 percent of the variation in the estimated training effects are explained by differences in training types, but the low p-level in Model 1 indicates that these differences are highly statistically significant.

Few of the other covariates appear to be significantly related to the training effect. The percentage of the variation explained by the covariates increases as more are added, but not greatly. Thus, for example, labor market conditions, as measured by the unemployment rate and the percentage of the workforce in manufacturing, appear to have little influence on the extent to which training programs increase the earnings of women. The coefficients on these variables are usually small and never statistically significant. As noted earlier, we also examined whether the maximum AFDC payment for which a family of three is eligible had any influence on the effect of training programs on women's earnings. The coefficient on this variable (which is not shown in Table 6), while positive, was relatively small and never approached conventional levels of statistical significance. Similarly, there is no evidence that more expensive training programs perform better for women than less expensive ones. However, the

Table 6
Variation in Program Effects, Mixed Effects Model Results
Women

	1	2	3	4	5	6	7	8
Training Type								
Classroom skills training	1787 ***	1702 ***	1672 ***	1594 ***	1518 ***	1306 ***	1295 ***	1335 ***
	(128)	(133)	(132)	(146)	(152)	(175)	(172)	(189)
Basic education	-211	-120	-172	-223	-176	-196	62	66
	(501)	(496)	(494)	(513)	(511)	(497)	(495)	(494)
CT+Basic ed	-302	-59	-374	-428	308	247	-122	-283
	(497)	(505)	(537)	(547)	(708)	(693)	(695)	(728)
OJT	1619 ***	1568 ***	1594 ***	1591 ***	1637 ***	1443 ***	1570 ***	1644 ***
	(241)	(240)	(238)	(242)	(243)	(252)	(253)	(325)
Subsidized work	816 ***	880 ***	1010 ***	1088 ***	1051 ***	811 ***	848 ***	784 ***
	(228)	(227)	(234)	(242)	(242)	(260)	(255)	(269)
Mix of classroom and workplace training	1410 ***	1430 ***	1436 ***	1490 ***	1479 ***	1331 ***	1580 ***	1538 ***
	(142)	(140)	(139)	(145)	(145)	(154)	(177)	(186)
Experimental dummy		-388 **	-104	-59	-43	-101	287	350
		(184)	(224)	(259)	(258)	(253)	(287)	(299)
White			539 **	537 **	353	244	-19	-33
			(248)	(249)	(272)	(270)	(281)	(283)
Non-white			381 *	676 **	430	253	216	277
			(230)	(343)	(374)	(373)	(365)	(376)
Unemployment rate				58	-1	11	-176	-236
				(199)	(202)	(196)	(204)	(220)
Unemployment rate squared				-8	-4	-3	6	9
				(13)	(13)	(13)	(13)	(14)
Percent manufacturing employment					42	37	-3	-7
					(26)	(25)	(29)	(29)
Years since training						181 **	96	97
						(80)	(84)	(84)
Year of training							-60 ***	-59 **
							(22)	(23)
Program cost								0.025
								(0.034)
Program cost missing								-82
								(489)
Percentage random variance (σ^2/v^*)	56.2%	55.3%	54.6%	54.9%	54.5%	52.7%	51.2%	51.1%
Percentage explained variance ($1-[v^*/s^2]$)	17.6%	19.7%	21.8%	22.1%	23.1%	26.4%	29.1%	30.1%
P-level for significance of covariates	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1. P-level given is from unweighted model.
* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

lack of a negative coefficient on the time since training variable suggests that, as for men, the effects of training persist for at least several years after the training is complete. In fact, there is weak evidence that they possibly increase, but this effect is statistically significant only in Model 6. In contrast to the results for men, there is also evidence that earlier training programs had larger effects than more recent programs.

For every model estimated for women, the random component is at least half of the total error variance. This is in contrast to the results for men, where the random component declines in importance as more covariates are added (recall that for men the random component declines to less than 20-percent of the total error variance). Thus, for women, a large part of the variation in training effects cannot be explained by sampling error or observable variables. However, as was shown earlier, there is less variation in the overall distribution of training effects for women than for either men and youth, so in a sense there is less to explain.

c. Youth

It was noted earlier that evidence from evaluations suggest that, overall, training programs for youths have a negligible effect on their earnings. As implied in Table 7, this is apparently because most estimates of earnings effects for youths pertain to a mix of classroom and workplace training. Such estimates account for over two-third of our observations and imply that training programs are ineffective for youths. However, Table 7 suggests that certain types of training may be effective for youths. Specifically, and similarly to women and to some extent men, classroom skills training seems to have a positive payoff, increasing earnings by at least $1,500 (Models 1 - 7) and possibly by as much as $2,000 (Model 8). Moreover, the coefficients on OJT and subsidized work are also positive, albeit usually fairly small and almost always statistically insignificant, suggesting that these programs could be somewhat effective.

As mentioned earlier, there is considerably more variation in the estimated training effects for

Table 7
Variation in Program Effects, Mixed Effects Model Results
Youth

	1	2	3	4	5	6	7	8
Training Type								
Classroom skills training	1547 ***	1524 ***	1430 ***	1581 ***	1563 ***	1634 ***	1599 ***	1954 ***
	(445)	(455)	(387)	(355)	(360)	(362)	(382)	(377)
CT+Basic ed	-377	-431	-537	-124	-141	-86	-115	-1006 *
	(834)	(862)	(638)	(521)	(521)	(518)	(540)	(498)
OJT	277	251	172	344	332	405	366	709
	(584)	(593)	(545)	(524)	(527)	(529)	(543)	(539)
Subsidized work	75	58	-60	69	48	319	423	273
	(305)	(313)	(254)	(231)	(344)	(392)	(446)	(417)
Mix of classroom and workplace	-323	-327	-335 *	-206	-207	-26	-19	-91
training	(199)	(200)	(167)	(152)	(151)	(195)	(198)	(186)
Female	69	77	92	-202	-194	-237	-261	-247
	(250)	(253)	(210)	(203)	(219)	(220)	(227)	(212)
Males and females	201	217	173	-45	-29	-77	-225	-658
	(391)	(398)	(321)	(293)	(313)	(313)	(448)	(425)
Experimental dummy		68	29	-40	-10	-48	-197	186
		(243)	(243)	(221)	(288)	(288)	(413)	(412)
White			-1070 ***	-1147 ***	-1136 ***	-1017 ***	-993 ***	-1279 ***
			(305)	(284)	(313)	(322)	(335)	(321)
Non-white			470	-239	-221	-404	-396	-458
			(288)	(478)	(497)	(512)	(520)	(527)
Unemployment rate				-232 ***	-230 ***	-219 ***	-229 ***	-255 ***
				(67)	(68)	(68)	(72)	(68)
Unemployment rate squared				5 ***	5 ***	5 ***	5 ***	5 ***
				(1)	(1)	(1)	(1)	(1)
Percent manufacturing employment					2	-8	0	-78 *
					(40)	(40)	(46)	(47)
Years since training						-140	-130	-116
						(96)	(101)	(92)
Year of training							16	-13
							(37)	(38)
Program cost								0.108 ***
								(0.042)
Program cost missing								969 ***
								(249)
Percentage random variance (σ^2/v^*)	24.8%	25.1%	14.6%	9.1%	8.4%	8.2%	8.7%	5.8%
Percentage explained variance ($1-[v^*/s^2]$)	11.7%	11.6%	23.1%	28.3%	29.2%	29.6%	29.5%	32.3%
P-level for significance of covariates	0.034	0.054	0.001	0.000	0.000	0.000	0.000	0.001

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1. P-level given is from unweighted model.
* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

youth than for mean and women. Thus, as indicated in Table 7, several covariates in addition to training type significantly contribute to explaining this variation. For example, as in the case of men, training appears less effective for whites than for non-whites or for the omitted mixed groups of whites and non-whites. Moreover, the coefficients for female youths and for the mixed group of male and female youths are generally negative in the more elaborate model specifications, albeit not statistically significant, suggesting that the payoff from training may be larger for male youths (the omitted group).

One reason government funded training programs for youths are often found to be ineffective is because youth unemployment is usually relatively high. The findings in Table 7 suggest that training effects for youth are highly sensitive to the unemployment rate, an effect that is very robust to model specification. Training becomes increasingly less effective as unemployment increases, but then becomes more effective, a pattern that is similar for men. However, the upswing doesn't occur until an unemployment rate of about 45 percent, which is the maximum unemployment rate in the sample. In Model 4, for example, at the mean unemployment rate in the sample (22.3 percent), the average youth enrolled in classroom skills training experienced an increase of $1,581 in annual earnings one year after the training took place. But at an unemployment rate of 30 percent, the effect is only $86, and it becomes negative after that.

Not only are the effects of government training programs on the earnings of youths often found to be negligible, or even detrimental, the negative (although statistically insignificant) coefficient on the time since training variable suggest that whatever positive effects occur tend to diminish over time. The findings in Table 7 also indicate that there is no evidence that more recent training programs for youths are any better than older programs.

In contrast to the findings for men and women, training effects on the earnings of youths appear to increase with program cost. For example, the results in Model 8 imply that a $1,000 increase in program cost increases the effect of training by about $108. This effect is statistically significant at the 1-percent level.

Finally, the random component is smaller for youth than for men and women, but as in the case of men, the random component declines in importance as additional covariates are added to the model. In Model 1, the random component comprises about 40 percent of the total error variance, while in Model 8, it comprises only about 18 percent of the total error variance. Overall, the contribution of the covariates is statistically significant in every model and they explain between 10 and 34 percent of the total variation in the estimated training effects.

7. Sensitivity of Results

It is of interest to determine whether the results are sensitive to extreme values of the training effects. Table 8 reports estimates of Model 8 when the highest and lowest three training effect estimates are excluded from the samples. For comparison purposes, the full model results for Model 8 are also reported in Table 8.

For the most part, the results are fairly robust to excluding the extreme values. Although the magnitude of the coefficients change somewhat, their signs and levels of statistical significance are generally similar. However, a few of the coefficients change considerably, particularly for youth. For example, the coefficient for youth on "mix of classroom and workplace training" changes from -91 (not statistically significant) to -543 (statistically significant) and the coefficient on "nonwhite" changes from -458 (not statistically significant) to 2,754 (statistically significant).

8. Conclusions

This report has presented findings from an exploratory meta-analysis of evaluations of voluntary government training programs for the disadvantaged. Exploiting the fact that there are numerous estimates of program effects on earnings and that there is considerable variation among them, we examined the influence of a number of factors on the size of the earnings effects estimates. Men, women, and youths were studied separately.

Table 8
Sensitivity Tests for Model 8

	Men		Women		Youth	
	Full Sample	Reduced Sample	Full Sample	Reduced Sample	Full Sample	Reduced Sample
Training Type						
Classroom skills training	131	82	1335 ***	1272 ***	1954 ***	769 **
	(404)	(363)	(189)	(163)	(377)	(322)
Basic education	--	--	66	373	--	--
	--	--	(494)	(496)	--	--
CT+Basic ed	--	--	-283	373	-1006 *	--
	--	--	(728)	(652)	(498)	--
OJT	312	258	1644 ***	1593 ***	709	-585
	(436)	(407)	(325)	(288)	(539)	(539)
Subsidized work	-961	126	784 ***	1230 ***	273	277
	(1267)	(1126)	(269)	(242)	(417)	(324)
Mix of classroom and workplace training	465	984 ***	1538 ***	1450 ***	-91	-543 ***
	(413)	(384)	(186)	(160)	(186)	(148)
Female	--	--	--	--	-247	85
	--	--	--	--	(212)	(166)
Males and females	--	--	--	--	-658	-610 *
	--	--	--	--	(425)	(327)
Experimental dummy	1218 ***	950 *	350	-78	186	-615 *
	(616)	(559)	(299)	(267)	(412)	(347)
White	-684	-937 ***	-33	-3	-1279 ***	-702 ***
	(430)	(366)	(283)	(248)	(321)	(180)
Non-white	-326	-419	277	243	-458	2754 ***
	(630)	(535)	(376)	(327)	(527)	(405)
Unemployment rate	111	194	-236	-37	-255 ***	-125 **
	(437)	(372)	(220)	(189)	(068)	(050)
Unemployment rate squared	-8	-18	9	-2	5 ***	0.12
	(25)	(21)	(14)	(12)	(001)	(001)
Percent manufacturing employment	384 ***	224 *	-7	25	-78 *	86 **
	(122)	(118)	(29)	(26)	(047)	(034)
Years since training	111	-4	97	45	-116	47
	(130)	(112)	(84)	(73)	(092)	(070)
Year of training	91	-5	-59 **	-38 *	-13	131 ***
	(74)	(73)	(23)	(20)	(038)	(030)
Program cost	-0.004	-0.085	0.025	-0.006	0.108 ***	0.139 ***
	(0.135)	(0.120)	(.034)	(.031)	(.042)	(.037)
Program cost missing	976	982 *	-82	-77	969 ***	-263
	(647)	(594)	(489)	(431)	(249)	(201)
Sample Size	83	77	133	127	99	93

Note: Reduced sample excludes the three highest and three lowest impact estimates.
Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean.
Years since training is centered around 1. P-level given is from unweighted model.
* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

It is apparent from the findings that the factors explaining variations in earnings estimates differ considerably among the three groups. This is not surprising because the training effects also differ greatly among the groups. On average, the effects tend to be largest for women, quite modest for men, and negligible for youths. It is also not surprising that our findings are the most robust for women, the group for whom we have the largest number of estimates of program earnings effects, and the least robust for men, the group for whom we have the fewest estimates.

Somewhat discouragingly, although the United States has over three decades of experience in running training programs, the evidence from our meta-analysis suggests that, in terms of improving the earnings of participants, they have not become more effective over time. The learning curve does indeed appear to be steep! Moreover, the effects are rarely found to be large. Even for adult women, who, according to program evaluations, are by far the most successful group of trainees, the vast majority of estimates indicate that training programs increase earnings by less than $2,000 a year for a typical trainee (see Figure 1).

A bit of good news from the meta-analysis is that, in contrast to frequent suggestions, we found no evidence that the effects of training on the earnings of adults diminish over time once participants leave a program. Indeed, there is some indication that earnings effects may even increase somewhat. However, we only observe the effects of training on earnings for a maximum of five years after the training is completed.

The findings suggest, that except for youths, a more expensive training program is not necessarily a superior one, in terms of increasing earnings. Nonetheless, it is possible that, for a given level of expenditure, certain types of training may be superior to others. Thus, we put considerable effort into developing a meaningful classification of the types of training represented by our sample of earnings effect estimates. Each measure of earnings effects was assigned to one of the following six categories: classroom skills training, basic education, mixed classroom skills training and basic education, OJT, subsidized work, and mixed classroom and workplace training.

None of these categories were found to be consistently superior to the other five. Although basic education was represented by relatively few observations in our sample, the findings we have suggest that it is ineffective. On the more positive side, classroom skills training was almost always found to be effective.

Looking separately at our three trainee groups, most types of training (with the possible exception of basic education) seem to be effective and work equally well for women. In contrast, only classroom skills training appears to be effective for youth. Firm conclusions cannot be drawn for men. Unfortunately, the findings for training type are simply not robust to specification changes for this group.

We found no evidence to support the hypothesis that a higher unemployment rate increases training program effects on earnings, except possibly at very high levels of unemployment. There is, however, some evidence that supports the contrary hypothesis, at least over the range of unemployment that we observe for our sample of earnings effect estimates. This evidence is especially strong for youths.

Perhaps surprisingly, government funded training seems less effective for white men and white youths than for non-white and racially mixed groups of men and youths. One possible explanation for this finding, which did not occur for women, is that white workers face fewer employment barriers and, hence, can more readily find jobs on their own without the aid of training.

Meta-analysis is usually performed on fairly small samples of earnings effect estimates and this study is no exception. However, compared to many previous studies our samples are fairly large. Nonetheless, at least for youths, our results are somewhat sensitive to the exclusion of extreme values and it is important for future meta-analyses of training programs to perform sensitivity tests of the findings.

Meta-analysis has the potential to synthesize systematically available information in a policy area of interest, and thus facilitate the accumulation of knowledge in the area. However, the ability to conduct useful meta-analyses of evaluations of voluntary government training programs is greatly limited by the lack of uniformity in current evaluation practices. Although it is important that evaluators not be overly constrained so that they have the opportunity to innovate, several fairly simple changes in future training program evaluations could considerably aid the cumulation of evidence concerning these programs.

First, and most simply, evaluation reports should provide exact standard errors or probability values (p-values) for program effect estimates so that the weights needed for meta-analyses can be readily computed. Although many evaluation reports do provide the necessary information, some do not, and, in these instances, it is necessary to impute standard errors. Sometimes the imputations can be subject to considerable error.

Second, more consistent reporting practices across evaluation studies would greatly facilitate meta-analysis, and synthesis in general. For example, it would be very helpful if all studies reported the same key population characteristics (e.g., the average age of the sample population, the percentage with a high school education, etc.).

Third, meta-analysis would be greatly facilitated if different training program evaluations estimated program effects for a similar set of subgroups. Possible candidates for separate analyses are white and non-white trainees and trainees with and without a high school degree, but there are certainly others as well.

Fourth, for purposes of the meta-analysis, we assigned training program effect estimates to six different categories. Unfortunately, two of these categories – mixed classroom skills training and basic education and mixed classroom and workplace training – are so broad that they are not especially useful for public policy purposes. Nonetheless, they were necessitate by the nature of the earnings effect estimates available from existing evaluation reports. The mixed classroom and workplace training category makes our findings for youths especially difficult to interpret, as it accounts for over two-thirds of our observations for this group. It would be useful for meta-analysis, as well as serve other purposes, if evaluators could develop and agree upon a training program typology and then estimate separate earnings effects for trainees falling into each of the resulting categories.

Table A-1
Mean Effects of Training by Program and Group

	Men			Women			Youth		
		Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Overall Mean	318 **	471 ***	337 ***	1417 ***	832 ***	1422 ***	-92	-28	-27
	(152)	(36)	(130)	(120)	(22)	(94)	(139)	(35)	(124)
Number of observations	83	83	83	133	133	133	99	99	
Mean by program									
MDTA	642 ***	577 ***	652 ***	2159 ***	1907 ***	2185 ***	n/a	n/a	n/a
	(236)	(42)	(217)	(176)	(46)	(152)			
SIME/DIME	-1805 *	-1891	-1881	525	470	480	n/a	n/a	n/a
	(960)	(1948)	(2085)	(489)	(829)	(908)			
SupportedWork	1310 *	705 *	1001	860 **	591 ***	742 **	28	137	117
	(679)	(375)	(703)	(399)	(171)	(356)	(607)	(174)	(388)
CETA	6	72	-79	1346 ***	1251 ***	1366 ***	698 **	49	393 *
	(248)	(74)	(225)	(159)	(48)	(145)	(335)	(47)	(238)
JOBS68	-149	932 ***	135	2069 ***	1878 ***	1934 ***	n/a	n/a	n/a
	(480)	(220)	(481)	(346)	(183)	(352)			
JTPA	761	785 **	770	1179 ***	887 ***	1065 ***	-357	87	-166
	(554)	(342)	(546)	(282)	(158)	(273)	(464)	(223)	(358)
FIP	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Massachusetts ET	n/a	n/a	n/a	501	81 **	565 *	n/a	n/a	n/a
				(370)	(33)	(308)			
Home Health Aide	n/a	n/a	n/a	1558 ***	902 ***	1344 ***	n/a	n/a	n/a
				(261)	(104)	(245)			
NJ OJT	n/a	n/a	n/a	1088	1092 ***	1089 *	n/a	n/a	n/a
				(692)	(301)	(601)			
Maine TOPS	n/a	n/a	n/a	1077	1213 **	1150	n/a	n/a	n/a
				(692)	(523)	(750)			
MFSP	n/a	n/a	n/a	309	-72	201	n/a	n/a	n/a
				(326)	(192)	(325)			
Job Corps	n/a	n/a	n/a	n/a	n/a	n/a	-221	-338	-283
							(289)	(97)	(223)
NYC/OS	n/a	n/a	n/a	n/a	n/a	n/a	-1055	-56	-495
							(402)	(99)	(311)
Jobstart	n/a	n/a	n/a	n/a	n/a	n/a	332	278	317
							(536)	(181)	(343)
New Chance	n/a	n/a	n/a	n/a	n/a	n/a	-308	-308	-308
							(1607)	(131)	(876)

Note: Standard erors in parentheses.
* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Table A-2
Variation in Program Effects
All Model Specifications
Men

	Specification 1			Specification 2			Specification 3			Specification 4		
		Weighted			Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Training Type												
Classroom skills training	495 **	568 ***	581 ***	515 **	645 ***	652 ***	512 **	794 ***	636 ***	157	562 ***	227
	(220)	(42)	(198)	(221)	(56)	(205)	(223)	(71)	(207)	(207)	(91)	(189)
OJT	359	786 ***	322	363	798 ***	308	375	780 ***	329	808 **	728 ***	598 **
	(368)	(183)	(359)	(369)	(184)	(361)	(371)	(184)	(364)	(324)	(185)	(301)
Subsidized work	-75	-161	-171	-194	-213	-387	-116	-353	-277	-14	-294	-193
	(475)	(183)	(460)	(496)	(185)	(483)	(513)	(218)	(512)	(441)	(219)	(411)
Mix of classroom and workplace	52	162 **	57	63	233 ***	107	18	250 ***	73	459	480 ***	530 **
training	(346)	(82)	(273)	(346)	(89)	(277)	(351)	(92)	(282)	(324)	(104)	(237)
Experimental dummy				396	549 **	755	716	414	975 *	1638 ***	630 **	1475 ***
				(467)	(270)	(479)	(578)	(290)	(558)	(519)	(295)	(467)
White							264	-429 **	205	-1125 **	-429 **	-775 *
							(510)	(192)	(465)	(500)	(211)	(426)
Non-white							537	-63	405	1416 ***	477 **	1217 ***
							(482)	(192)	(443)	(513)	(224)	(426)
Unemployment rate										-2102 ***	-29	-1131 ***
										(393)	(113)	(320)
Unemployment rate squared										132 ***	-11	61 ***
										(28)	(8)	(23)
Percent manufacturing employment												
Years since training												
Year of training												
Program cost												
Program cost missing												
Program												
SIME/DIME												
SupportedWork												
CETA												
JOBS68												

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1.

* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Table A-2
Variation in Program Effects
All Model Specifications
Men (concl.)

	Specification 5			Specification 6			Specification 7			Specification 8		
		Weighted			Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Training												
Classroom skills training	143	106	71	-8	135	26	128	129	24	221	27	131
	(185)	(108)	(157)	(254)	(137)	(206)	(248)	(137)	(194)	(472)	(308)	(404)
OJT	1072 ***	1081 ***	874 ***	895 **	1101 ***	828 ***	840 **	1104 ***	848 ***	803 *	391	312
	(295)	(191)	(256)	(360)	(199)	(290)	(345)	(199)	(274)	(458)	(361)	(436)
Subsidized work	-934 **	-677 ***	-771 **	-1204 **	-625 **	-856 *	-1228 **	-568 **	-848 **	-1506	241	-961
	(444)	(224)	(361)	(542)	(273)	(440)	(519)	(280)	(415)	(1438)	(937)	(1267)
Mix of classroom and workplace training	748 **	775 ***	766 ***	548	798 ***	717 ***	98	989 ***	582 *	58	1077 ***	465
	(296)	(110)	(187)	(376)	(130)	(237)	(394)	(245)	(336)	(448)	(307)	(413)
Experimental dummy	2393 ***	965 ***	1568 ***	2425 ***	959 ***	1583 ***	2105 ***	1167 ***	1342 ***	2194 ***	731	1218 **
	(493)	(298)	(394)	(495)	(299)	(396)	(488)	(375)	(460)	(677)	(466)	(616)
White	-1140 **	-879 ***	-851 **	-1067 **	-886 ***	-835 **	-556	-1041 ***	-700 *	-599	-995 ***	-684
	(446)	(219)	(338)	(455)	(220)	(341)	(473)	(277)	(400)	(507)	(281)	(430)
Non-white	-389	-510 **	-84	-787	-408	-222	-196	-447	-162	-194	-652	-326
	(610)	(256)	(453)	(764)	(397)	(612)	(761)	(399)	(590)	(795)	(410)	(630)
Unemployment rate	-1013 **	308 **	-168	-728	236	-71	-653	150	44	-690	309	111
	(427)	(121)	(293)	(539)	(246)	(414)	(516)	(263)	(402)	(549)	(271)	(437)
Unemployment rate squared	80 ***	-25 ***	7	66 **	-21	2	56 *	-16	-6	59 *	-24	-8
	(27)	(8)	(19)	(32)	(14)	(24)	(30)	(15)	(23)	(32)	(15)	(25)
Percent manufacturing employment	354 ***	238 ***	232 ***	388 ***	226 ***	247 ***	502 ***	166 **	288 ***	504 ***	234 ***	384 ***
	(79)	(30)	(59)	(88)	(48)	(73)	(94)	(81)	(107)	(100)	(91)	(122)
Years since training				153	-22	41	183	-55	58	182	-7	111
				(176)	(64)	(123)	(169)	(74)	(121)	(172)	(77)	(130)
Year of training							123 ***	-43	34	131 **	-11	91
							(44)	(47)	(59)	(55)	(58)	(74)
Program cost										0.032	-0.107	-0.004
										(0.156)	(0.102)	(0.135)
Program cost missing										140	1039 **	976
										(681)	(519)	(647)
Program												
SIME/DIME												
SupportedWork												
CETA												
JOBS68												

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1.

* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Table A-3
Variation in Program Effects
All Model Specifications
Women

	Specification 1			Specification 2			Specification 3			Specification 4		
		Weighted			Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Training												
Classroom skills training	1681 ***	1754 ***	1787 ***	1623 ***	1765 ***	1702 ***	1604 ***	1551 ***	1672 ***	1559 ***	1655 ***	1594 ***
	(142)	(42)	(128)	(144)	(49)	(133)	(143)	(51)	(132)	(151)	(59)	(146)
Basic education	-164	-294	-211	-52	-293	-120	-168	-87	-172	-258	219	-223
	(586)	(205)	(501)	(582)	(205)	(496)	(585)	(205)	(494)	(601)	(208)	(513)
CT+Basic ed	-224	-502 *	-302	13	-525 *	-59	-342	-1181 ***	-374	-415	-518 *	-428
	(507)	(273)	(497)	(517)	(278)	(505)	(556)	(282)	(537)	(564)	(295)	(547)
OJT	1711 ***	1672 ***	1619 ***	1652 ***	1678 ***	1568 ***	1657 ***	1698 ***	1594 ***	1682 ***	1517 ***	1591 ***
	(233)	(133)	(241)	(232)	(134)	(240)	(231)	(134)	(238)	(235)	(136)	(242)
Subsidized work	800 ***	970 ***	816 ***	838 ***	962 ***	880 ***	951 ***	1370 ***	1010 ***	1014 ***	1281 ***	1088 ***
	(246)	(113)	(228)	(244)	(114)	(227)	(249)	(117)	(234)	(259)	(120)	(242)
Mix of classroom and workplace	1488 ***	438 ***	1410 ***	1542 ***	449 ***	1430 ***	1561 ***	833 ***	1436 ***	1595 ***	894 ***	1490 ***
training	(162)	(27)	(142)	(163)	(36)	(140)	(163)	(42)	(139)	(167)	(44)	(145)
Experimental dummy				-376 *	36	-388 **	-80	350 ***	-104	0	-381 ***	-59
				(195)	(80)	(184)	(244)	(82)	(224)	(278)	(126)	(259)
White							511 *	970 ***	539 **	501 *	574 ***	537 **
							(275)	(63)	(248)	(276)	(83)	(249)
Non-white							422 *	999 ***	381 *	688 *	286 *	676 **
							(253)	(67)	(230)	(370)	(153)	(343)
Unemployment rate										-50	534 ***	58
										(228)	(60)	(199)
Unemployment rate squared										-1	-32 ***	-8
										(15)	(4)	(13)
Percent manufacturing employment												
Years since training												
Year of training												
Program cost												
Program cost missing												
Program												
SIME/DIME												
SupportedWork												
CETA												
JOBS68												
JTPA												
Massachusetts ET												
Home Health Aide												
NJ OJT												
Maine TOPS												
MFSP												

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1.

* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Table A-3
Variation in Program Effects
All Model Specifications
Women (concl.)

	Specification 5			Specification 6			Specification 7			Specification 8		
		Weighted			Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Training												
Classroom skills training	1506 ***	1444 ***	1518 ***	1259 ***	1345 ***	1306 ***	1209 ***	1305 ***	1295 ***	1338 ***	1358 ***	1335 ***
	(155)	(70)	(152)	(184)	(76)	(175)	(182)	(78)	(172)	(205)	(84)	(189)
Basic education	-206	267	-176	-236	239	-196	-41	347	62	-11	289	66
	(600)	(209)	(511)	(589)	(209)	(497)	(585)	(214)	(495)	(584)	(215)	(494)
CT+Basic ed	258	730 **	308	146	774 **	247	-20	522	-122	-351	216	-283
	(730)	(368)	(708)	(718)	(368)	(693)	(710)	(385)	(695)	(744)	(398)	(728)
OJT	1725 ***	1554 ***	1637 ***	1507 ***	1440 ***	1443 ***	1642 ***	1452 ***	1570 ***	1723 ***	1613 ***	1644 ***
	(236)	(136)	(243)	(249)	(140)	(252)	(252)	(140)	(253)	(314)	(192)	(325)
Subsidized work	978 ***	1244 ***	1051 ***	707 **	1104 ***	811 ***	787 ***	1083 ***	848 ***	587 *	1010 ***	784 ***
	(259)	(120)	(242)	(279)	(127)	(260)	(277)	(128)	(255)	(303)	(129)	(269)
Mix of classroom and workplace training	1586 ***	917 ***	1479 ***	1399 ***	856 ***	1331 ***	1615 ***	946 ***	1580 ***	1517 ***	891 ***	1538 ***
	(166)	(44)	(145)	(181)	(48)	(154)	(202)	(63)	(177)	(211)	(65)	(186)
Experimental dummy	33	-412 ***	-43	-23	-450 ***	-101	329	-328 **	287	475	-245 *	350
	(278)	(126)	(258)	(274)	(126)	(253)	(311)	(138)	(287)	(326)	(140)	(299)
White	339	250 **	353	243	145	244	44	39	-19	-17	-22	-33
	(297)	(100)	(272)	(294)	(105)	(270)	(302)	(115)	(281)	(305)	(117)	(283)
Non-white	427	-26	430	222	-152	253	276	-199	216	364	4	277
	(411)	(163)	(374)	(412)	(167)	(373)	(406)	(168)	(365)	(416)	(179)	(376)
Unemployment rate	-100	462 ***	-1	-101	455 ***	11	-298	414 ***	-176	-446 *	255 ***	-236
	(229)	(62)	(202)	(225)	(62)	(196)	(238)	(64)	(204)	(255)	(79)	(220)
Unemployment rate squared	4	-27 ***	-4	5	-26 ***	-3	16	-24 ***	6	23	-17 ***	9
	(15)	(4)	(13)	(15)	(4)	(13)	(15)	(4)	(13)	(16)	(5)	(14)
Percent manufacturing employment	40	63 ***	42	34	63 ***	37	10	41 ***	-3	-1	36 **	-7
	(28)	(11)	(26)	(27)	(11)	(25)	(29)	(15)	(29)	(30)	(15)	(29)
Years since training				221 **	92 ***	181 **	147	74 ***	96	150	67 **	97
				(93)	(28)	(80)	(97)	(29)	(84)	(97)	(29)	(84)
Year of training							-50 **	-25 **	-60 ***	-46 **	-18	-59 **
							(22)	(11)	(22)	(23)	(11)	(23)
Program cost										0.058	0.055 ***	0.025
										(0.036)	(0.016)	(0.034)
Program cost missing										43	-140	-82
										(482)	(278)	(489)
Program												
SIME/DIME												
SupportedWork												
CETA												
JOBS68												
JTPA												
Massachusetts ET												
Home Health Aide												
NJ OJT												
Maine TOPS												
MFSP												

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1.

* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Table A-4
Variation in Program Effects
All Model Specifications
Youth

	Specification 1			Specification 2			Specification 3			Specification 4		
		Weighted			Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Training												
Classroom skills training	1341 *	1338 ***	1547 ***	1401 *	1145 ***	1524 ***	1396 *	1128 ***	1430 ***	1605 ***	1262 ***	1581 ***
	(633)	(270)	(445)	(645)	(280)	(455)	(604)	(280)	(387)	(593)	(283)	(355)
CT+Basic ed	-717	-338 *	-377	-566	-644 ***	-431	-606	-650 ***	-537	-71	-381 *	-124
	(1629)	(181)	(834)	(1660)	(215)	(862)	(1553)	(215)	(638)	(1522)	(227)	(521)
OJT	383	200	277	443	28	251	438	47	172	647	171	344
	(633)	(480)	(584)	(645)	(484)	(593)	(604)	(484)	(545)	(593)	(486)	(524)
Subsidized work	106	-155	75	148	-159	58	84	-127	-60	215	-14	69
	(478)	(137)	(305)	(486)	(137)	(313)	(456)	(141)	(254)	(462)	(145)	(231)
Mix of classroom and workplace training	-648 *	-215 *	-323	-656 *	-258 *	-327	-618 *	-296 ***	-335 *	-413	-225 *	-206
	(268)	(105)	(199)	(270)	(106)	(200)	(253)	(108)	(167)	(257)	(110)	(152)
Female	409	30	69	394	127	77	425	145	92	47	-36	-202
	(343)	(126)	(250)	(346)	(131)	(253)	(324)	(131)	(210)	(343)	(143)	(203)
Males and females	334	224	201	292	306 *	217	-79	386 *	173	-295	124	-45
	(677)	(139)	(391)	(684)	(143)	(398)	(664)	(153)	(321)	(699)	(172)	(293)
Experimental dummy				-198	305 ***	68	-731 *	407 ***	29	-871 *	305 *	-40
				(370)	(114)	(243)	(431)	(136)	(243)	(464)	(138)	(221)
White							-1621 ***	-348 *	-1070 ***	-1776 ***	-510 ***	-1147 ***
							(474)	(166)	(305)	(535)	(177)	(284)
Non-white							-137	670 ***	470	-1381 *	354	-239
							(464)	(162)	(288)	(730)	(349)	(478)
Unemployment rate										-247 *	-185 ***	-232 ***
										(132)	(44)	(67)
Unemployment rate squared										5 *	3 ***	5 ***
										(2)	(1)	(1)
Percent manufacturing employment												
Years since training												
Year of training												
Program cost												
Program cost missing												
Program												
SupportedWork												
CETA												
JTPA												
Job Corps												
NYC/OS												
Jobstart												

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1.

* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

Table A-4
Variation in Program Effects
All Model Specifications
Youth (concl.)

	Specification 5			Specification 6			Specification 7			Specification 8		
		Weighted			Weighted			Weighted			Weighted	
	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects	Unweighted	Fixed Effects	Mixed Effects
Training												
Classroom skills training	1736 ***	1147 ***	1563 ***	1794 ***	1234 ***	1634 ***	1762 ***	1190 ***	1599 ***	2008 ***	1804 ***	1954 ***
	(605)	(288)	(360)	(607)	(295)	(362)	(659)	(302)	(382)	(659)	(330)	(377)
CT+Basic ed	150	-561 *	-141	160	-476 *	-86	129	-514 *	-115	-886	-1189 ***	-1006 *
	(1534)	(243)	(521)	(1533)	(250)	(518)	(1561)	(256)	(540)	(1625)	(282)	(498)
OJT	778	61	332	837	138	405	805	93	366	938	623	709
	(605)	(489)	(527)	(607)	(491)	(529)	(659)	(496)	(543)	(652)	(510)	(539)
Subsidized work	692	-378 *	48	918	-155	319	930	-32	423	698	32	273
	(640)	(227)	(344)	(674)	(273)	(392)	(685)	(331)	(446)	(685)	(332)	(417)
Mix of classroom and workplace training	-429 *	-270 *	-207	-203	-128	-26	-200	-114	-19	-160	-193	-91
	(257)	(112)	(151)	(334)	(149)	(195)	(337)	(150)	(198)	(334)	(154)	(186)
Female	-104	89	-194	-133	20	-237	-137	3	-261	-97	-156	-247
	(370)	(155)	(219)	(371)	(162)	(220)	(375)	(164)	(227)	(369)	(170)	(212)
Males and females	-478	318	-29	-460	243	-77	-514	94	-225	-793	-524	-658
	(719)	(195)	(313)	(719)	(202)	(313)	(833)	(305)	(448)	(834)	(322)	(425)
Experimental dummy	-1214 *	563 ***	-10	-1203 *	497 ***	-48	-1257 *	356	-197	-1162	570 *	186
	(562)	(185)	(288)	(562)	(191)	(288)	(705)	(288)	(413)	(774)	(303)	(412)
White	-1421 *	-732 ***	-1136 ***	-1325 *	-648 ***	-1017 ***	-1315 *	-597 ***	-993 ***	-1309 *	-1031 ***	-1279 ***
	(628)	(207)	(313)	(634)	(215)	(322)	(642)	(228)	(335)	(644)	(241)	(321)
Non-white	-964	102	-221	-1095	-90	-404	-1104	-49	-396	-1029	-289	-458
	(826)	(369)	(497)	(834)	(392)	(512)	(842)	(397)	(520)	(870)	(436)	(527)
Unemployment rate	-277 *	-160 ***	-230 ***	-251 *	-159 ***	-219 ***	-253 *	-165 ***	-229 ***	-273 *	-236 ***	-255 ***
	(135)	(46)	(68)	(137)	(46)	(68)	(138)	(47)	(72)	(140)	(51)	(68)
Unemployment rate squared	6 *	3 ***	5 ***	6 *	3 ***	5 ***	6 *	3 ***	5 ***	6 *	4 ***	5 ***
	(2)	(1)	(1)	(2)	(1)	(1)	(2)	(1)	(1)	(2)	(1)	(1)
Percent manufacturing employment	-85	58 *	2	-93	47	-8	-85	55 *	0	-97	-66 *	-78 *
	(79)	(28)	(40)	(79)	(29)	(40)	(100)	(32)	(46)	(103)	(38)	(47)
Years since training				-187	-97	-140	-179	-92	-130	-182	-62	-116
				(177)	(67)	(96)	(188)	(67)	(101)	(185)	(68)	(92)
Year of training							9	17	16	56	-29	-13
							(67)	(26)	(37)	(78)	(28)	(38)
Program cost										0.171 *	0.098 ***	0.108 ***
										(0.083)	(0.035)	(0.042)
Program cost missing										471	1124 ***	969 ***
										(461)	(191)	(249)
Program												
SupportedWork												
CETA												
JTPA												
Job Corps												
NYC/OS												
Jobstart												

Note: Standard erors in parentheses. All variables except training types and years since training are centered around the sample mean. Years since training is centered around 1.

* Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

References

I. Studies Used in the Meta-Analysis

Ashenfelter, Orley. "Estimating the Effects of Training Programs on Earnings," *Review of Economics and Statistics*, February 1978, 60(1), pp. 47-57.

Ashenfelter, Orley and Card, David. "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs," *Review of Economics and Statistics*, November 1985, 67(4), pp. 648-60.

Auspos, Patricia, Cave, George, and Long. David, *Maine: Final Report on the Training Opportunities in the Private Sector Program*, New York. Manpower Demonstration Research Corporation, 1988.

Bassi, Laurie J. "The Effect of CETA on the Postprogram Earnings of Participants," *Journal of Human Resources*, Fall 1983, 18(4), pp. 539-56.

___. "Estimating the Effect of Training Programs with Non-Random Selection," *Review of Economics and Statistics* 66(1), February 1984, pp. 36-43.

Bell, Stephen H. and Orr, Larry L. "Is Subsidized Employment Cost Effective for Welfare Recipients?: Experimental Evidence from Seven State Demonstrations," *Journal of Human Resources,* Spring 1994, 19(1) , pp. 42-61.

Bloom, Howard S. "Estimating the Effect of Job Training Programs Using Longitudinal Data: Ashenfelter's Findings Reconsidered," *Journal of Human Resources*, Fall, 1984b, 19(4), pp. 544-56.

___."What Works for Whom?: CETA Impacts for Adult Participants," *Evaluation Review*, August 1987, 11(4), pp. 510-27.

Bryant, Edward C. and Rupp, Kalman. "Evaluating the Impact of CETA on Participant Earnings," *Evaluation Review,* August 1987, 11(4), pp. 473-92.

Burghardt, John et al. *Evaluation of the Minority Female Single Parent Demonstration*, Volume I, Summary Report. New York: The Rockefeller Foundation, October 1992.

Cain, Glen G. "Benefit-Cost Estimates for Job Corps," Discussion Paper No. 9-68, Institute for Research on Poverty, University of Wisconsin, Madison, September 1968.

Cave, George et al. *JOBSTART: Final Report on a Program for School Dropouts*, New York, Manpower Demonstration Research Corporation, October 1993.

Cooley, Thomas M.; McGuire, Timothy W. and Prescott, Edward C. "Earnings and Employment Dynamics of Manpower Trainees: An Econometric Analysis," in *Research in Labor Economics.* Ed: Farrell E. Bloch. Greenwich: JAI Press, 1979.

Couch, Kenneth. "New Evidence on the Long-Term Effects of Employment and Training Programs," *Journal of Labor Economics*, October 1992, 10(4), pp. 380-88.

Dickinson, Katherine P. and Richard W. West. "Impacts of Counseling and Education Subsidy Programs," *Final Report of the Seattle-Denver Income Experiment Volume I: Design and Report Part IV,* SRI International, 1983.

Dickinson, Katherine P.; Johnson, Terry R. and West, Richard W. *An Analysis of the Impact of CETA Programs on Components of Earnings and Other Outcomes*. Final Report prepared for the U.S. Department of Labor, Employment and Training Administration, Menlo Park, CA: SRI International, November 1984.

___."An Analysis of the Impact of CETA Programs on Participants' Earnings," *Journal of Human Resources*, Winter 1986, 21(1), pp. 64-91.

___. "The Impact of CETA Programs on Components of Participants' Earnings," *Industrial and Labor Relations Review,* April 1987a, 40(3), pp. 430-40.

___. "An Analysis of the Sensitivity of Quasi-Experimental Net Impact Estimates of CETA Programs," *Evaluation Review,* August 1987b, 11(4), pp. 452-72.

Freedman, Stephen; Bryant, Jan and Cave, George. *New Jersey: Final Report on the Grant Diversion Project*. New York: Manpower Demonstration Research Corporation, 1988.

Gay, Robert and Borus, Michael. "Validating Performance Indicators for Employment and Training Programs," *Journal of Human Resources,* Winter 1980, 15(1), pp. 29-48.

Hollister, Robinson G.; Kemper, Peter and Maynard, Rebecca, eds. *The National Supported Work Demonstration*. Madison: University of Wisconsin Press, 1984.

Kiefer, Nicholas M. "The Economic Benefits From Four Government Training Programs," in *Research in Labor Economics*, Supplement 1. Ed.: Farrell E. Bloch, ed. 1979, pp. 159-86.

___ . "Federally Subsidized Occupational Training and the Employment and Earnings of Male Trainees." *Journal of Econometrics,* 1978, 8(1), pp. 111-25.

Mallar, Charles et al. *Evaluation of the Economic Impact of the Job Corps Program: Third Follow-Up Report*. Princeton: Mathematica Policy Research, September 1982.

Orr, Larry L. et al. *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*. Washington, D.C.: The Urban Institute Press, 1996.

Nightingale-Smith, Demetra et al. *Evaluation of the Massachusetts Employment and Training (ET) Program.* Urban Institute Report 91-1. Washington, D.C.: The Urban Institute Press, 1991.

Quint, Janice C. et al. *New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Young Mothers and Their Children*. New York: Manpower Demonstration Research Corporation, September 1994.

Schochet, Peter Z., John Burghardt, and Steven Glazerman. *National Job Corps Study: The Short-Term Impacts of Job Corps on Participants' Employment and Related Outcomes Final Report*, Mathematica Policy Research, Inc., Princeton, NJ, February 2000.

Westat. *Continuous Longitudinal Manpower Survey: Summary of Net Impact Results*. Report MEL 84-02 prepared for the U.S. Department of Labor under contract No. 23-24-75-07. Rockville, MD: Westat, April 1984.

Zambrowski, Amy and Gordon, Anne. *Evaluation of the Minority Female Single Parent Demonstration: Fifth-year impacts at CET*. Princeton: Mathematica Policy Research, Inc., December 1993.

II. Other References

Barnow, Burt S. "The Impact of CETA Programs on Earnings: A Review of the Literature," *Journal of Human Resources*, Spring 1987, 22(2), pp. 157-93.

Bloom, Howard S., Carolyn J. Hill, and James Riccio. *Modeling the Performance of Welfare-to-Work Programs: The Effects of Program Management and Services, Economic Environment, and Client Characteristics*, unpublished manuscript, New York: Manpower Demonstration Research Corporation, May 2001.

Cohen, Jacob. *Statistical Power Analysis in the Social Sciences*, Hillsdale, NJ, Erlbaum, 1988.

Cook, Thomas. et al. *Meta-Analysis for Explanation: A Casebook*, New York: Russell Sage Foundation, 1992.

Cooper, Harris. M. and Larry. V. Hedges (eds.). *The Handbook of Research Synthesis*, New York: The Russell Sage Foundation, 1994.

Finifter, David H. "An Approach to Estimating Net Earnings Impact of Federally Subsidized Employment and Training Programs," *Evaluation Review*, August 1987, 11(4), pp. 528-47.

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. "Evaluating Government Programs for the Economically Disadvantaged," *Journal of Economic Literature*, Vol. 35, No. 4, December 17, 1997, pp. 1,809-1,855.

Greenberg, David, Robert Meyer, Charles Michalopoulos, and Michael Wiseman. *Explaining Variation in the Effects of Welfare-to-Work Programs*, Institute for Research on Poverty Discussion Paper No. 1225-01, Madison, WI: University of Wisconsin, March 2001.

Greenberg, David and Michael Wiseman. "What Did the OBRA Demonstrations Do?" in *Evaluation Design for Welfare and Training Programs* (I. Garfinkel and C. Manski, eds.), Cambridge Mass.: Harvard University Press, 1992.

Greenberg, David and Michael Wiseman. *Work-Welfare Initiatives: How Do We Add Things Up?* unpublished manuscript, October 1990.

Heckman, James, Robert LaLonde, and Jeffrey Smith. "The Economics and Econometrics of Active Labor Market Programs," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Volume 3A, Amsterdam: North Holland, 1999, pp. 1,865-2,097.

Hedges, Larry V. "Advances in Statistical Methods for Meta-Analysis," in *Issues in Data Synthesis*, San Francisco: Jossey-Bass, 1984, pp. 25-42.

Hedges, Larry V., R. D. Laine, and R. Greenwald. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher*, Vol. 23, No. 3, 1994, pp. 5-14.

Hedges, Larry V. “Meta-Analysis,” *Journal of Educational Statistics*, 1992.

Hunt, Morton. *How Science takes Stock: The Story of Meta-Analysis*, New York: Russell Sage Foundation, 1997.

Hunter, J. and F. Schmidt. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, Newbury Park, CA: Sage Publications, 1990.

Jarrell, S. B. and T. D. Stanley. “A Meta-Analysis of the Union-Nonunion Wage Gap,” *Industrial and Labor Relations Review*, Vol. 40, January 1990, pp. 54-67.

LaLonde, Robert J. “The Promise of Public Sector-Sponsored Training Programs. *Journal of Economic Perspectives*, Spring 1995, 9(2), pp. 149-68.

Lipsey, Mark W. and David B. Wilson. *Practical Meta-Analysis*, Thousand Oaks, CA: Sage 2001.

Raudenbush, Stephen W. “Random Effects Models.” Chapter 20 in Cooper, Harris N. and Larry V. Hedges, eds., *The Handbook of Research Synthesis*, New York, Russell Sage Foundation, 1992, pp. 301-321.

Rosenthal, Robert. *Meta-Analytic Procedures for Social Research* (revised edition). Newbury Park, CA: Sage Publications, 1991.