Cornell University
ILR School

ILR Review

Volume 57 | Number 1 Article 2

October 2003

A Meta-Analysis of Government-Sponsored Training Programs

David H. Greenberg
University of Maryland

Charles Michalopoulos
MDRC

Philip K. Robins
University of Miami

A Meta-Analysis of Government-Sponsored Training Programs

Abstract

This study uses meta-analysis to synthesize findings from 31 evaluations of 15 voluntary government-funded training programs for the disadvantaged that operated between 1964 and 1998. On average, the earnings effects of the evaluated programs seem to have been largest for women, quite modest for men, and negligible for youths. For men and women, the earnings effects of training appear to have persisted for at least several years after the training was complete. Classroom skills training was apparently effective in increasing earnings, but basic education was not. There is no evidence that more expensive training programs performed better than less expensive ones. Although the United States has more than three decades of experience in running training programs, the programs do not appear to have become more effective over time.

Cover Page Footnote

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. A previous version of this paper was presented in a seminar at MDRC and at the 2001 annual meeting of the Association for Public Policy and Management in Washington, D.C. The authors acknowledge the valuable comments of Burt Barnow, Howard Bloom, Marvin Mandell, James Riccio, and especially Jeffrey Smith.

A META-ANALYSIS OF GOVERNMENT-SPONSORED TRAINING PROGRAMS

DAVID H. GREENBERG, CHARLES MICHALOPOULOS, and PHILIP K. ROBINS*

This study uses meta-analysis to synthesize findings from 31 evaluations of 15 voluntary government-funded training programs for the disadvantaged that operated between 1964 and 1998. On average, the earnings effects of the evaluated programs seem to have been largest for women, quite modest for men, and negligible for youths. For men and women, the earnings effects of training appear to have persisted for at least several years after the training was complete. Classroom skills training was apparently effective in increasing earnings, but basic education was not. There is no evidence that more expensive training programs performed better than less expensive ones. Although the United States has more than three decades of experience in running training programs, the programs do not appear to have become more effective over time.

Since the 1960s, federal and state governments have funded training programs designed to increase the earnings of low-income individuals who have ended their formal education. These programs have been envisioned as tools for combating unemployment and poverty and, more recently, as a tool for decreasing transfer payments by increasing the earnings of recipients of government transfers.

Although many evaluations of these programs have been published, there have been few attempts to formally synthesize the results. Studies of an individual program typically focus on whether the program "works" by, for example, increasing the earnings of those who participate. There have also been several recent summaries of these programs, but they have focused more on the overall effectiveness of the programs than on factors that make one program more effective than another (for example, see LaLonde 1995; Friedlander, Greenberg, and Robins 1997; Heckman, LaLonde, and Smith 1999). In addition, the summary studies have rarely used formal statistical tools to take the analysis beyond simple pattern recognition—essentially, visual in-

*David H. Greenberg is Professor Emeritus, Department of Economics, University of Maryland; Charles Michalopoulos is Senior Research Associate, MDRC; and Philip K. Robins is Professor, Department of Economics, University of Miami. 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. A previous version of this paper was presented in a seminar at MDRC and at the 2001 annual meeting of the Association for Public Policy and Management in Washington, D.C. The authors acknowledge the valuable comments of Burt Barnow, Howard Bloom, Marvin Mandell, James Riccio, and especially Jeffrey Smith.

A copy of the STATA data file used in the analysis is available from the third author, Philip K. Robins, at Department of Economics, P.O. Box 248126, Coral Gables, FL 33124; probins@miami.edu.

Industrial and Labor Relations Review, Vol. 57, No. 1 (October 2003). © by Cornell University.
0019-7939/00/5701 $01.00

spection of the data to spot possible links between program effects and program characteristics.

This paper uses meta-analysis, a statistical tool for synthesizing research findings across evaluations, to systematically analyze the findings of 31 studies investigating the effectiveness of 15 voluntary government training programs for the disadvantaged that operated in the United States between 1962 and 1998. Meta-analysis consists of procedures for extracting findings and other information from empirical research studies, assembling this information into a data base, and then analyzing the data using modified versions of standard statistical methods.[1] Until recently, meta-analysis has rarely been applied to economic issues, but its procedures are well developed through extensive use in numerous other areas of research, including medicine, psychology, education, and criminal justice. Descriptions of recent applications in those and other contexts can be found in Jarrell (1990), Hedges, Laine, and Greenwald (1994), Hunt (1997), and several chapters of Cook et al. (1992).

The objective of the analysis in this paper is to increase knowledge about the types of government training programs for the disadvantaged that are most effective. Specifically, we address the following questions:

—Do voluntary training programs increase the earnings of adult men, adult women, and youths?

—Do more expensive training programs have larger effects than less expensive programs?

—Do effects on earnings vary with the services provided? For example, does classroom training appear to be more or less effective than on-the-job training?

—Are the programs' effects larger for some demographic groups than for others?

—Are the programs' effects influenced by labor market conditions?

—Do the effects of the programs grow or decline over time after participants leave the programs?

—Do the findings of studies based on random assignment differ from the findings of studies based on nonexperimental designs?

—Have training programs conducted in recent years been more successful than earlier training programs?

Data

The Sample of Studies

Our sample includes every evaluation of voluntary U.S. government-funded training programs we could locate that (1) was conducted after 1974, (2) used individual data, and (3) compared program and comparison groups to determine the program's effects on earnings.[2] While the 1975 start date is admittedly somewhat arbitrary, it was selected for several reasons. First, the program evaluations published during the relatively short period demarcated by the 1975 start date are manageable in number for a meta-analysis. Second, 1975 is sufficiently early to capture several evaluations of programs conducted under the Manpower Development and Training Act (MDTA), the first major post–World War II national training program in the United States. Third, methods used in evaluating training programs became more sophisticated after 1974 (see Friedlander, Greenberg, and Robins 1997:1827–29). For example, 1975 marked the beginning of

[1]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).

[2]We excluded studies that based their estimates on a pre–post comparison of trainees because there is considerable evidence that such comparisons are unreliable due to the likelihood that factors other than the training cause the earnings of trainees to change over time (see, for example, Moffitt 1991). We also excluded secondary analyses of training programs that did not have as their central goal the estimation of the effects of training programs—for example, studies using data from random assignment evaluations to test whether various nonexperimental methodologies can successfully overcome selection problems (for example, LaLonde 1986; Fraker and Maynard 1987; Friedlander and Robins 1995).

the National Supported Work Demonstration, the first major random assignment evaluation of a training program.[3] In all, usable information from 31 studies of 15 voluntary training programs met these criteria. Because two of the studies (Kiefer 1978; Gay and Borus 1980) each evaluate four programs, we have 37 separate evaluations of 15 programs. 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 provided separate effect estimates for program participants at different locations. As a result, the total number of effect estimates from the studies is much larger than 37.

Table 1 provides summary information on the training programs and evaluations included in the meta-analysis.[4] The table indicates whether each program 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, the major evaluation studies of the program, and the evaluation method used (experimental or nonexperimental).[5]

The Dependent Variable

This paper analyzes the programs' effects on earnings. Training programs may, of course, affect other outcomes, such as employment, welfare and unemployment compensation payments, crime rates, and feelings of satisfaction. We focus on earnings because a major objective of government-sponsored training programs is to increase earnings of participants. Furthermore, all 31 studies estimated effects on earnings, while effects on other outcomes were estimated less frequently. Since earnings are measured in dollars, results are readily pooled across studies, although there are issues concerning appropriate adjustments for inflation over time and cost of living differences across study sites. The Gross Domestic Product (GDP) chain-type price index, published by the U.S. Department of Commerce, is used to adjust all estimates of training program effects on earnings to 1999 dollars.

As will be discussed below, one technique used in meta-analysis relies on the variance of each estimate to weight the studies. Many evaluations of training programs report exact measures of the 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%, 5%, or 10% level 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.

The imputations assumed that the p-value of each estimate was located at the midpoint of the possible range. If the level of statistical significance was reported to exceed the 5% level but not the 1% level, for example, it was assumed that the p-value equaled .03 (the midpoint between .01 and .05). Similarly, if the level of statistical significance exceeded 1%, it was assumed that p = .005 (the midpoint between zero and .01); if the level of statistical significance exceeded 10% but not 5%, it was assumed that p = .075 (the midpoint between .05 and .1); and, finally, if the earn-

[3]We include one pre-1974 study (Cain 1968) because it is a frequently cited early evaluation of the Job Corps program.

[4]One of these programs, the Seattle-Denver Income Maintenance Experiments (SIME/DIME), is mainly remembered as testing a negative income tax program. However, SIME/DIME also randomly assigned a subset of experimental subjects to a program with counseling and training subsidies. It is the effects for that component of SIME/DIME that are included in the present meta-analysis.

[5]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*[a]	*Evaluation Study*	*Method of Evaluation*
Manpower Development and Training Act (MDTA)	National	1962–73	Disadvantaged adults and youths	Classroom Training, On-the-Job Training	Ashenfelter (1978), Cooley et al. (1979), Kiefer (1978, 1979), Gay and Borus (1980),Bloom (1984)	Nonexperimental
Neighborhood Youth Corps (NYC)	National	1964–73	Disadvantaged youths	Classroom Training, Paid Work Experience	Kiefer (1979), Gay and Borus (1980)	Nonexperimental
Job Opportunities in the Business Sector (JOBS68)	Demonstration[b]	1967–73	Disadvantaged adults	On-the-Job Training	Kiefer (1979), Gay and Borus (1980)	Nonexperimental
Job Corps	National	1964–present	Disadvantaged youths	Classroom Training, Paid Work Experience	Cain (1968), Gay and Borus (1980), Kiefer (1979), Mallar et al. (1982), Schochet, Burghardt, and Glazerman (2000)	Nonexperimental Experimental
Seattle-Denver Income Maintenance Experiment (SIME/DIME)	Demonstration	1971–78	Low-income adults	Classroom Training	Dickinson and West (1983)	Experimental
Comprehensive Employment and Training Act (CETA)	National	1973–83	Disadvantaged adults and youths	Classroom Training, On-the-Job Training, Paid Work Experience, Public Service Employment	Westat (1984), Ashenfelter and Card (1985), Bassi (1983, 1984), Bloom (1987), Bryant and Rupp (1987), Dickinson et al. (1984, 1986, 1987a, 1987b)	Nonexperimental
National Supported Work Demonstration	Demonstration	1975–78	Long-term AFDC recipients, ex-addicts, ex-offenders, high school dropouts	Paid Work Experience with training	Hollister, Kemper, and Maynard (1984), Couch (1992)	Experimental
AFDC Homemaker–Home Health Aide Demonstration	Demonstration	1983–86	AFDC recipients	Paid Work Experience with training	Bell and Orr (1994)	Experimental
Maine Training Opportunities in the Private Sector (TOPS)	Demonstration	1983–86	AFDC recipients	On-the-Job Training, Unpaid Work Experience	Auspos, Cave, and Long (1988)	Experimental
New Jersey Grant Diversion Project	Demonstration	1984–87	AFDC recipients	On-the-Job Training	Freedman, Bryant, and Cave (1988)	Experimental
Minority Female Single Parent Demonstration (MFSP)	Demonstration	1982–87	Low-income minority single mothers	Classroom Training, On-the-Job Training	Burghardt et al. (1992), Zambrowski et al. (1983)	Experimental
Massachusetts Employment and Training Choices (ET)	Demonstration[c]	1983–89	AFDC recipients	Classroom Training, Unpaid Work Experience[d]	Nightingale et al. (1991)	Nonexperimental
Jobstart Demonstration	Demonstration	1985–88	High school dropouts	Classroom Training	Cave et al. (1993)	Experimental
New Chance Demonstration	Demonstration	1989–92	AFDC high school dropouts	Classroom Training, Paid Work Experience, Unpaid Work Experience	Quint et al. (1994)	Experimental
Job Training Partnership Act (JTPA)	National	1983–98	Disadvantaged adults and youths	Classroom Training, On-the-Job Training	Orr et al. (1996)	Experimental

[a]Most programs with training components also provided assistance with job search.
[b]A national program, but classified as a demonstration because it was short-lived and featured only a single training activity.
[c]A state-run version of a national program, but classified as a demonstration because its research interest lies mainly in the large scale of its voluntary approach to training for welfare recipients.
[d]Other services were also provided, including job development and college assistance.

ings effect estimate was not statistically significant, it was assumed that p = .3 (the midpoint between .1 and .5).[6]

Because of the crucial role of the standard errors in conducting a meta-analysis and because standard errors were imputed in almost one-quarter of the cases, it is important to determine whether the results are sensitive to the method of imputation. In a later section of this paper ("Sensitivity Tests"), we present results from an analysis of the sensitivity of the estimates to the method of imputation.

Explanatory Variables

We use a number of variables to try to explain the variation in programs' effects. With the exception of information on cost, data to construct the explanatory variables were available for all of the programs listed in Table 1. Following the lead of most training program evaluations, we performed separate analyses for adult men, adult women, and youths.

Program characteristics. The first program characteristic we consider is the type of training that was offered. Training types varied across programs, but consisted of one or more of the following: remedial education, classroom vocational or skills training, on-the-job training (OJT) in private sector jobs, and subsidized employment in the public and nonprofit sectors. Most of the programs also offered structured job search, which was typically, but not always, combined with one or more of the other program components.

In characterizing training type, we distinguished between classroom and workplace training. We then defined three categories of classroom training—basic education, classroom skills training, and a combination of basic education and skills training ("CT + basic ed")—and two categories of workplace training—on-the-job training (OJT) and subsidized work.[7] We had no prior expectations about which types of training would be most effective in increasing earnings.

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 and to learn about the relationship between cost and effectiveness, we constructed a variable that measures the program cost per participant.[8] We anticipated that, all else equal, more costly programs would be more successful ones. Unfortunately, program cost per participant is unavailable for adults for the Job Opportunities in the Business Sector (JOBS68) program and for youths for both the CETA and Neighborhood Youth Corps (NYC) programs. 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. The studies listed in Table 1 do not provide a consistent set of measures of the characteristics of program enrollees (for example, average age, average educational achievement, and marital status). However, many of the studies estimated separate effects for white and minority group enrollees. In addition, the

[6]Standard errors were imputed for 71 cases (23% of the estimated earnings effects). Of these, 20 were statistically significant at the 1% level, 8 were significant at the 5% level, none were significant at the 10% level, and 43 were not statistically significant at the 10% level or higher.

[7]OJT positions are permanent positions provided by private sector employers who receive subsidies from the training agency during the training period. The trainees receive regular wages. In subsidized work (sometimes called "paid work experience"), trainees are placed in temporary positions at government and non-profit agencies and receive a stipend.

[8]Where it was available (mainly experimental studies), we used administrative cost per participant net of training costs for comparison group members. In other cases (mainly nonexperimental studies), we used gross costs. Training costs for comparison group members were probably small in most nonexperimental studies, however. For some kinds of training (primarily paid work experience), costs include payments to participants.

studies of youth enrollees usually estimated separate effects for boys and girls. Thus, we constructed one set of three dummy variables indicating whether each estimate pertains to white trainees, nonwhite trainees, or a mixed group of both white and nonwhite trainees, and a second set indicating whether each estimate for youths pertains to boys, girls, or boys and girls combined.

We did not have prior expectations as to whether training programs for youths are more effective for boys or for girls. As for the relationship between race and program effects on earnings, two plausible opposing hypotheses can be formulated. On the one hand, since white training program enrollees tend to have higher levels of formal education, tend to live in neighborhoods that are more accessible to jobs, and tend to be less subject to discrimination than program enrollees from nonwhite groups, we might expect training to have the largest effects for white enrollees, mid-range effects for racially mixed groups, and the smallest effects for minority enrollees. On the other hand, white workers may be better able to succeed in the labor market on their own than nonwhite workers. If so, training will have the largest effects for nonwhites, mid-range effects for racially mixed groups, and the smallest effects for whites.

Area economic conditions. The effects of training are likely to be influenced by economic conditions. We use three variables to investigate the influence of economic conditions: the unemployment rate, the percentage of the work force in manufacturing, and (for women only) the maximum Aid to Families with Dependent Children (AFDC) payment for a family of three, each measured at the time and place of the corresponding program earnings effect estimate.

The influence of the unemployment rate on the earnings effects of training is theoretically ambiguous. On the one hand, trainees may enjoy a competitive advantage over similar non-trainees when the unemployment rate is high and jobs are difficult to find. If so, the relationship between the unemployment rate and the earnings effect will be positive. On the other hand, training may only be helpful if unemployment is low and jobs are readily available for trainees. In that case, the relationship will be negative.

We expected the percentage of the work force in manufacturing to be positively related to the effects of training programs on earnings. Traditionally, manufacturing jobs pay low-skilled workers higher wages than do jobs in the service industry. Thus, a proportionately high number of jobs in manufacturing should serve both to motivate low-skilled people to seek training and employment and to reward them with higher earnings once they find a job.

We obtained data on unemployment rates and the percentage of the work force in manufacturing from U.S. government publications such as the *Monthly Labor Review*, *Employment and Earnings*, the *Economic Report of the President*, and *County Data Patterns,* as well as from the U.S. Bureau of Labor Statistics' web page. Many of the program effect estimates used in the meta-analysis pertain to trainees in two or more places. In such instances, the unemployment rate and the percentage of the work force in manufacturing were obtained for each site and a weighted average was computed, with the weights being the percentage of the evaluation sample contributed by each site.

Because each estimate we use is for a particular age and gender group, and often also for a specific racial group, we use an unemployment rate that is also age-, gender-, and race-specific. Because local unemployment rates are usually only reported for the entire local labor force, age-, gender-, and race-specific local unemployment rates are obtained by multiplying each reported local unemployment rate by the ratio of the national unemployment rate for the group of interest to the overall national unemployment rate.

AFDC payment levels might influence a program's effects on earnings for women. To control for this, we use the maximum AFDC payment level 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.

The anticipated relationship between the AFDC payment level and the effects of training is ambiguous. On the one hand, AFDC is a potential alternative to work for single parents. Generous AFDC payments may therefore reduce the incentive of some adult women to seek training and jobs and, consequently, 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. Some of the evaluations are from random assignment studies. Nonexperimental methods generally produced the largest negative and positive estimates, suggesting that the use of nonexperimental methods can result in substantial estimation error. To control for this possibility, we use a dummy variable that is set equal to one for random assignment studies, and zero for nonexperimental studies.[9] We did not have prior expectations about whether estimates from random assignment studies would be larger or smaller than estimates from other studies.

Years since training. Most of the studies tracked effects for two or three years after the training was received. A few studies measured effects during the fourth and even the fifth post-training year.[10] To examine how the effects of training change over time, we include a variable equal to the number of years since the training occurred.[11]

This "years since training" variable is used to test two opposite hypotheses. The first is that training gives workers a competitive advantage that diminishes over time as similar workers who did not receive the training catch up. The second is that training opens doors that allow participants to obtain additional training and other forms of human capital after they leave the program and take a job, in which case the program's effects will grow over time.

Calendar year of training. The earliest training in our studies occurred in 1964. Although training programs have offered the same basic services since then, program coordinators may have learned better ways to provide these services. To test this possibility, we use a simple time trend. If government-funded training programs have improved over time, there should be an upward trend in their effects. However, this variable may be subject to a downward bias if training has become increasingly available over time to members of the comparison groups used to obtain estimates of program effects on earnings.

Descriptive Statistics of the Sample

Table 2 shows the number of estimates for each of the 31 studies. Many of the studies produced estimates for different population groups (men, women, youths, blacks, whites, welfare recipients, non-welfare recipients, and so on) 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 youths.

All but four of the studies produced more than one estimate. Kiefer (1979) produced the most estimates, with 48 (15% of the

[9]In principle, it would be desirable to distinguish among different types of nonexperimental procedures. Preliminary attempts to do so, however, proved largely uninformative, primarily because of the relatively small sample sizes and difficulties in characterizing the substantial number of different nonexperimental procedures used.

[10]Estimates beyond the third year after training are available for the MDTA program, the Supported Work Program, and the MFSP program. Overall, 27% of our observations are for more than two years after training and 8% are for more than three years after training.

[11]For a more comprehensive analysis of the relationship between earnings effect estimates and years since training, see Greenberg, Robins, and Michalopoulos (forthcoming).

Table 2. Number of Estimates by Study.

Study	*Men*	*Women*	*Youths*	*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

sample), while the average number of estimates per study is just over 10 (315/31). Seventeen studies produced estimates for men, averaging just under 5 estimates per study (83/17); 23 studies produced estimates for women, averaging just under 6 estimates per study (133/23); and 15 studies produced estimates for youths, averaging just over 6.5 estimates per study (99/15).

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

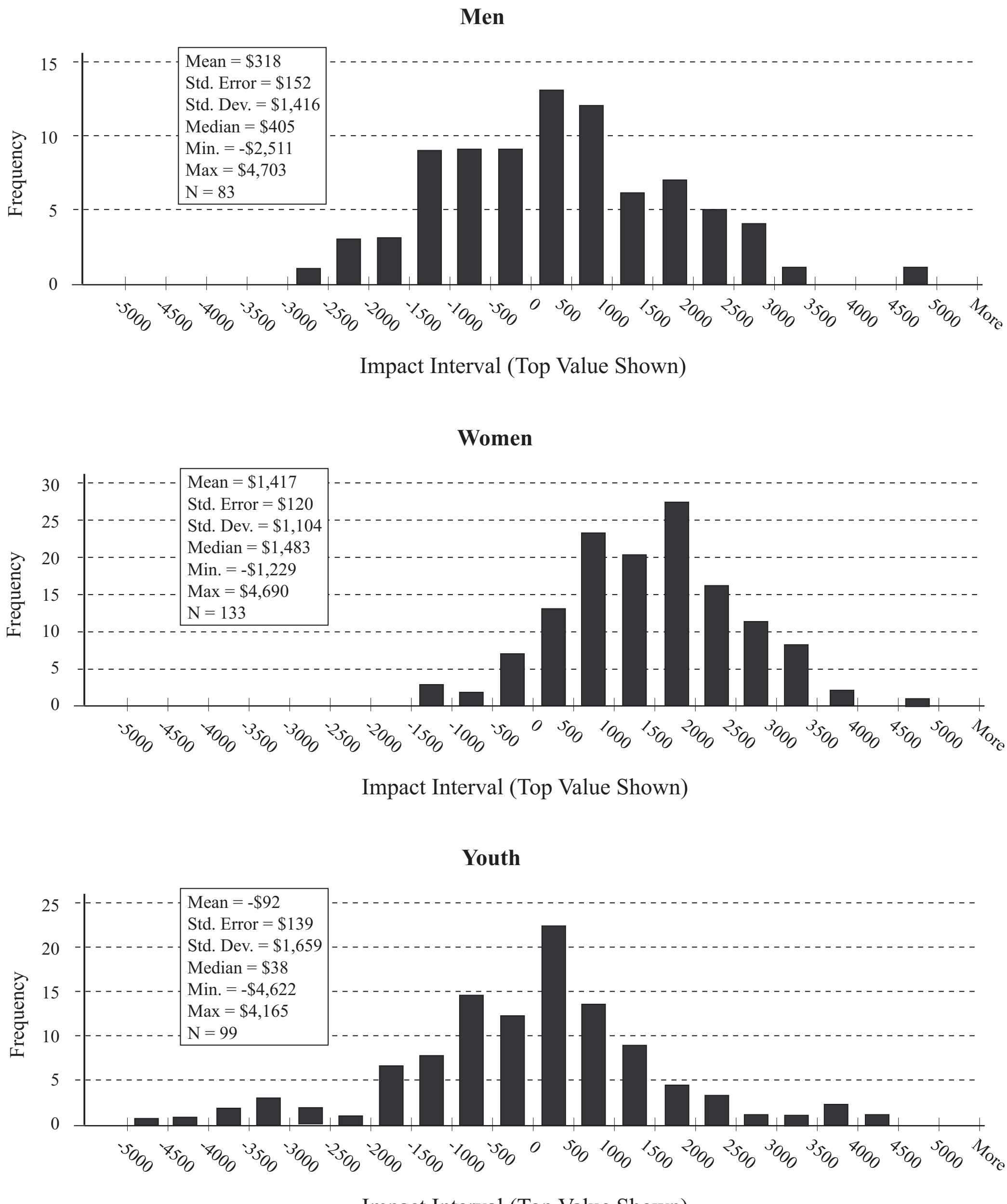
Among the three groups, women have by far the highest mean training effect: $1,417. This effect is significantly different from zero at the 1% level. Men have a mean effect of $318, which is significantly different from zero at the 5% level. Youths have a negative mean effect of –$92, which is not significantly different from zero.

Not only is training most effective for women, but the distribution of 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 youths, with a standard deviation of $1,659 and a range from –$4,622 to $4,165. Thus, while training appears to have been ineffective for youths, great uncertainty is associated with this conclusion.

Table 3 presents the mean earnings effect for each program, along with the proportion of the sample in each program. For men, the estimated effects vary from a low of –$1,805 for SIME/DIME to a high of $1,310 for Supported Work (significantly different from zero at the 10% level in both cases). Just over three-quarters of the estimates are for the MDTA and CETA programs, with MDTA having an average effect of $642 (significantly different from zero at the 1% level) and CETA an average effect of $6 (not significantly different from zero). For women, the estimates vary from a low of $309 for MFSP (not significantly different from zero) to a high of $2,159 for MDTA (significantly different from zero at the 1% 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% level. For youths, the estimates vary from a low of –$1,055 for NYC (significantly different from zero at the 1% level) to a high of $698 for CETA (significantly different from zero at the 5% level). About a fourth of the estimates for youths are for the CETA program and about a third are for the Job Corps program.[12]

[12]Because many current policy-makers place little weight on results from older programs like CETA and MDTA, it is of interest to see whether the results and

Figure 1. Histograms of Training Effects by Group.



conclusions of this paper would change if findings from these older programs were treated as less important than findings from more recent programs like JTPA. In an appendix available from the authors, we present results from analyses that put less weight on the CETA and MDTA programs relative to the other programs. Although some differences occur, the basic conclusions of the paper do not change.

The F-test that is reported at the bottom of Table 3 indicates that, for all three groups, the hypothesis that the effects are the same for each program can be rejected.[13] The hypothesis is only marginally rejected for youths (p = .062), moderately rejected for men (p = .044), and strongly rejected for women (p = .000). These tests imply that some of the variation in the effects is due to differences among the programs.

Statistical Model

To explore which measured factors are associated with the estimated training program effects, we specify the following statistical model, drawn from Raudenbush (1994):

$$T_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_p X_p + e_i + u_i, \quad (1)$$

where T_i is the estimated effect of training, the X's are observed characteristics of the studies that cause variation in the true program effects (such as type of training and demographic characteristics of the sample), the β's are coefficients to be estimated, e_i is sampling error (with variance v_i), and u_i is error due to unmeasured factors (with variance σ^2).[14] In other words, v_i arises because of within-study variation due to sampling error and u_i arises because of between-study variation due to unmeasured factors.

A standard result in the meta-analysis literature (see, for example, Chapter 8 of Hedges and Olkin 1985) is that the estimated parameters in equation (1) will have the smallest variance (under appropriate assumptions) when an observation (in our case, the estimated effect of a training program) receives a weight inversely proportional to its variance. With this weighting scheme, more precisely estimated observations receive greater weight than less precisely estimated observations. This makes intuitive sense, because the "true" population mean is probably close to the precise estimates, but could be quite far from the imprecise estimates.[15] This choice of weights does not affect the point estimates of the coefficients (they are consistent regardless of whether no weight is used or whether they are weighted by the inverse of their variances), but it does affect their level of statistical significance and can be thought of as a correction for heteroskedasticity.

After accounting for the measured factors (that is, the X's in the model given by equation 1), program effect estimates differ from one another for two reasons: (1) sampling error (e_i) and (2) unmeasured factors (u_i).[16] If there are no other unob-

[13]As discussed in some detail in the next section, in formal meta-analysis, weighted regressions are usually estimated to take account of the fact that the individual estimates of the effects of training programs are subject to sampling error. However, the F-tests reported in Table 3 are based on unweighted OLS regressions, which use the earnings effect estimates as the dependent variable, and dummy variables representing the programs listed in Table 3 as explanatory variables. Hence, the results in Table 3 do not account for sampling error. However, the same conclusions hold for tests based on the appropriately weighted regressions, which do account for sampling error.

[14]Theoretically, it might be possible to collect information on many of the factors that seem unmeasurable. However, studies of training programs do not produce a consistent set of information on program practices and other hard-to-measure factors. The most ambitious attempt to incorporate information on such factors is described in Bloom, Hill, and Riccio (2001).

[15]The simplest example involves the mean of two independent program effect estimates T_1 and T_2 with variances v_1 and v_2. The variance of the weighted mean program effect estimate, $pT_1 + (1-p)T_2$, is $p^2 v_1 + (1-p)^2 v_2$, which is minimized by choosing $p = (1/v_1)/(1/v_1 + 1/v_2)$.

[16]Two additional possible sources of variation are not taken into account in this specification: correlation between estimates for the same study in different years, and correlation among different studies of the same program using the same underlying data set (for example, many of the CETA studies). To some extent, we control for such variation through variables measuring training type, calendar time, and time since training. An alternative approach would be to calculate robust standard errors allowing for clusters in the data. We experimented with calculating robust standard errors using clusters for year of training. Although in several instances the standard errors were reduced, they typically rose by between 10% and 20%.

Table 3. Unweighted Mean Effects of Training by Program and Group. (Standard Errors in Parentheses)

	Men		*Women*		*Youths*	
Description	*Mean*	*Fraction of Sample*	*Mean*	*Fraction of Sample*	*Mean*	*Fraction of Sample*
Overall Mean	318** (152)	1.00	1,417*** (120)	1.00	–92 (139)	1.00
Mean by Program						
MDTA	642*** (236)	0.40	2,159*** (176)	0.23	—	—
SIME/DIME	–1805* (960)	0.02	525 (489)	0.03	—	—
Supported Work	1,310* (679)	0.05	860** (399)	0.05	28 (607)	0.07
CETA	6 (248)	0.36	1,346*** (159)	0.29	698** (335)	0.23
JOBS68	–149 (480)	0.10	2,069*** (346)	0.06	—	—
JTPA	761 (554)	0.07	1,179*** (282)	0.09	–357 (464)	0.12
Massachusetts ET	—	—	501 (370)	0.05	—	—
Home Health Aide	—	—	1,558*** (261)	0.11	—	—
New Jersey Grant Diversion Project	—	—	1,088 (692)	0.02	—	—
Maine TOPS	—	—	1,077 (692)	0.02	—	—
MFSP	—	—	309 (326)	0.07	—	—
Job Corps	—	—	—	—	–221 (289)	0.31
NYC	—	—	—	—	–1,055*** (402)	0.16
Jobstart	—	—	—	—	332 (536)	0.09
New Chance	—	—	—	—	–308 (1607)	0.01
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 errors in parentheses.
*Significant at 10% level. **Significant at 5% level. ***Significant at 1% level.

served factors other than sampling error, then u_i is identically zero for all studies and the weight used in estimating the model is the inverse of the sampling variance $(1/v_i)$, which is estimated as the inverse of the square of the standard error of the estimate. In this case, differences occur across estimates because of factors that can be measured, and because estimates were obtained from a sample rather than the population. The meta-analysis literature refers to this case as a "fixed effects" model. Using the estimated variances from each study produces a weighted mean estimate of \$471 for men (compared to an unweighted mean of \$318), \$832 for women (compared to \$1,417), and –\$28 for youths (compared to –\$92). Thus, except for women, the weighted means are close to the unweighted means. The weighted mean is significantly different from zero at the 1% level for men and women and is not statistically significant for youths.

It is easy to think of reasons why u_i might not be zero for all studies, since many features of the programs were unmeasured. One program might have accepted only the most able applicants, while another might have accepted all eligible applicants, for example. One classroom training program might have focused on literacy, another on math skills. One program might have been run by a charismatic leader, another by a bureaucratic administrator. All of these factors result in more variation in program effect estimates than can be explained by sampling error alone.

The meta-analysis literature refers to the case when u_i is not identically zero as a "mixed effects" model.[17] In the mixed effects model, the variance of the unexplained portion of the program effect estimate is $v_i + \sigma^2$. It can be shown (see Raudenbush 1994) that the weight minimizing the variance of the estimates of β is the inverse of this variance ($1/[v_i + \sigma^2]$). Clearly, the fixed effects model is a special case of the mixed effects model in which σ^2 is zero. It is therefore possible to test statistically the null hypothesis that the fixed effects model is consistent with the observed variation in program effect estimates across studies. This test is described below.

To estimate σ^2 in the mixed effects model, we use the method of moments estimator described by Raudenbush (1994). This estimator involves the following steps. Equation (1) is first estimated using ordinary least squares (OLS). The mean square residual variance from the regression is then used to calculate an estimate of σ^2, based on the formula

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

where MSR is the mean square residual from the OLS regression and k is a constant given by the formula (see Raudenbush 1994:319)

$$k = \Sigma v_i - \mathrm{trace}[\mathbf{XVX}(\mathbf{XX})^{-1}], \tag{3}$$

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

The mixed effects estimate of the mean program effect is \$337 for men (compared to a fixed effects mean of \$471), \$1,422 for women (compared to \$832), and –\$27 for youths (compared to –\$28). Thus, for men and women, the mixed effects means differ from the fixed effects means, but are similar to the unweighted means. For youths, the mixed effects mean is virtually identical to the fixed effects mean. The mixed effects mean is significantly different from zero at the 1% level for men and women, but is not significantly different from zero for youths.

For completeness, we estimated the fixed effects and mixed effects models as well as the unweighted OLS model. Using the test suggested by Raudenbush (1994:314), the unweighted estimates and fixed effects

[17] Sometimes the "mixed effects" model is referred to as a "random effects" model, because it includes the random term u_i, which captures random unobserved factors. This is somewhat of a misnomer, because the sampling error e_i is also assumed to be random. Nonetheless, throughout the remainder of this article, we refer to u_i as the "random component" of the error term.

model were emphatically rejected in favor of the mixed effects model for almost every model specification for all three groups.[18] We therefore present only the mixed effects results in the tables that follow. However, the reader should be aware that the test favoring the mixed effects model is based on the assumption that the standard errors derived from the studies are accurate. In an appendix available from the authors, we present the full set of estimates for all three models.

Results

Table 4 presents the means and standard deviations of the covariates defined earlier. Most of the earnings estimates—70% of those for men, 76% of those for women, and 79% of those for youths—are for either nonwhite trainees or a mixed group of white and nonwhite trainees. In addition, most of the estimates are nonexperimental (86% for men, 63% for women, and 69% for youths). The average unweighted unemployment rate in the study sites was 5.5% for men, 6.9% for women, and 22.3% for youths. Just under one-quarter of the work force 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 unweighted average program cost (in 1999 dollars) was $7,080 for men, $6,591 for women, and $8,782 for youths. 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 (youths).

To assess the robustness of key parameter estimates, we introduce the covariates sequentially in the following order: (1) training type, (2) whether the evaluation was a randomized experiment, (3) race and ethnicity, (4) the unemployment rate and its square, (5) the percentage 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 (8) program cost.[19] 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 one.

Tables 5–7 present the mixed effects model results for men, women, and youths, 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 $\sigma^2 + v$),[20] 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. The p-level is based on the unweighted OLS regression.

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

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

[19]An additional specification was estimated that included dummy variables for each of the programs. Due to high collinearity with the other covariates, however, the coefficients were very imprecisely estimated and the results were largely uninformative. In an appendix available from the authors, unweighted, fixed, and mixed effects estimates of a model including only the program variables are presented. 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.

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

Table 4. Unweighted Means and Standard Deviations of Covariates.

Variable	*Men (N = 83)*	*Women (N = 133)*	*Youths (N = 99)*
Training Type			
Classroom Skills Training (CT)	0.51 (0.50)	0.38 (0.49)	0.07 (0.26)
Basic Education	0.00 (0.00)	0.02 (0.15)	0.00 (0.00)
CT+Basic Education	0.00 (0.00)	0.03 (0.17)	0.01 (0.10)
OJT	0.18 (0.39)	0.14 (0.35)	0.07 (0.26)
Subsidized Work	0.11 (0.31)	0.13 (0.34)	0.16 (0.37)
Mix of Classrooom & Workplace Training	0.20 (0.41)	0.29 (0.46)	0.69 (0.47)
1 = Female Youths	0.00 (0.00)	0.00 (0.00)	0.47 (0.50)
1 = Male and Female Youths Combined	0.00 (0.00)	0.00 (0.00)	0.08 (0.27)
1 = Experimental Dummy	0.14 (0.35)	0.37 (0.48)	0.31 (0.47)
1 = White[a]	0.30 (0.46)	0.24 (0.43)	0.21 (0.41)
1 = Nonwhite[a]	0.40 (0.49)	0.31 (0.46)	0.23 (0.42)
Unemployment Rate (%)	5.55 (2.47)	6.95 (2.47)	22.27 (9.28)
Unemployment Rate Squared	36.80 (31.80)	54.35 (37.62)	581.29 (518.51)
Percent Manufacturing Employment	24.77 (3.31)	23.07 (4.97)	21.04 (4.10)
Years Since Training	2.17 (1.05)	1.99 (1.00)	2.01 (0.93)
Year of Training (1964 = 0)	7.90 (6.60)	11.76 (7.64)	12.37 (7.46)
Program Cost[b]	7,080 (3,573)	6,591 (3,690)	8,782 (4,031)

Variable	*Men (N = 83)*	*Women (N = 133)*	*Youths (N = 99)*
1 = Program Cost Missing	0.10 (0.30)	0.06 (0.24)	0.42 (0.50)
Program[c]			
1 = MDTA	0.40 (0.49)	0.23 (0.42)	0.00 (0.00)
1 = SIME/DIME	0.02 (0.15)	0.03 (0.17)	0.00 (0.00)
1 = Supported Work	0.05 (0.22)	0.05 (0.21)	0.07 (0.26)
1 = CETA	0.36 (0.48)	0.29 (0.45)	0.23 (0.42)
1 = JOBS68	0.10 (0.30)	0.06 (0.24)	0.00 (0.00)
1 = JTPA	0.07 (0.26)	0.09 (0.29)	0.12 (0.33)
1 = Massachusetts ET	0.00 (0.00)	0.05 (0.22)	0.00 (0.00)
1 = Home Health Aide	0.00 (0.00)	0.11 (0.31)	0.00 (0.00)
1 = NJ OJT	0.00 (0.00)	0.02 (0.12)	0.00 (0.00)
1 = Maine TOPS	0.00 (0.00)	0.02 (0.12)	0.00 (0.00)
1 = MFSP	0.00 (0.00)	0.07 (0.25)	0.00 (0.00)
1 = Job Corps	0.00 (0.00)	0.00 (0.00)	0.31 (0.47)
1 = NYC/OS	0.00 (0.00)	0.00 (0.00)	0.16 (0.37)
1 = Jobstart	0.00 (0.00)	0.00 (0.00)	0.09 (0.29)
1 = New Chance	0.00 (0.00)	0.00 (0.00)	0.01 (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.

on the training types can be interpreted as mean program effects for the average person in each type of training, a year after the training took place. 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 youths pertain to a mix of classroom and workplace training.

Men

Of the findings for the three groups we examined, those for men tend to be the

Table 5. Variation in Program Effects for Men: Mixed Effects Model Results. (Standard Errors in Parentheses)

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Training Type								
Classroom Skills Training	581*** (198)	652*** (205)	636*** (207)	227 (189)	71 (157)	26 (206)	24 (194)	131 (404)
OJT	322 (359)	308 (361)	329 (364)	598** (301)	874*** (256)	828*** (290)	848*** (274)	312 (436)
Subsidized Work	–171 (460)	–387 (483)	–277 (512)	–193 (411)	–771** (361)	–856* (440)	–848** (415)	–961 (1,267)
Mix of Classroom and Workplace Training	57 (273)	107 (277)	73 (282)	530** (237)	766*** (187)	717*** (237)	582* (336)	465 (413)
Experimental Dummy		755 (479)	975* (558)	1,475*** (467)	1,568*** (394)	1,583*** (396)	1,342*** (460)	1,218** (616)
White			205 (465)	–775* (426)	–851** (338)	–835** (341)	–700* (400)	–684 (430)
Non-White			405 (443)	1,217*** (426)	–84 (453)	–222 (612)	–162 (590)	–326 (630)
Unemployment Rate				–1,131*** (320)	–168 (293)	–71 (414)	44 (402)	111 (437)
Unemployment Rate Squared				61*** (23)	7 (19)	2 (24)	–6 (23)	–8 (25)
Percent Manufacturing Employment					232*** (59)	247*** (73)	288*** (107)	384*** (122)
Years Since Training						41 (123)	58 (121)	111 (130)
Year of Training							34 (59)	91 (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 ($1-[v^*/\sigma^2]$)	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

Notes: 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.

least robust and the most difficult to interpret, possibly because the number of observations for men is also the smallest. As a result, we cannot draw firm conclusions about the effectiveness of particular types of training for men. In Model 1, for instance, the effect of classroom skills training is $581. As indicated by the high p-level, however, this effect is not significantly different from the effect of the other training types. Moreover, the coefficients on individual training types change substantially as additional covariates are added to the model. Because none of the coefficients on training types in Table 5 exceeds $1,000, it seems unlikely that any type of training resulted in large positive effects on the annual earnings of men.

A number of interesting findings emerge from examining the coefficients on the remaining covariates reported in Table 5. Random assignment studies produced considerably larger earnings effects estimates for men than nonexperimental studies did. This could mean that nonexperimental studies have systematically understated the effects of training for men, or that random assignment was conducted for only the most effective programs.

Most specifications in Table 5 imply that white men benefited less from training than did the mixed and (probably) nonwhite groups. The effect of training for men appears to have been considerably higher in areas where there was more manufacturing employment. Model 8 implies, for example, that a one percentage point increase in manufacturing employment increased the effect of training by $384. This effect is very robust with respect to model specification.

The effects of training do not appear to have varied systematically with the unemployment rate, the cost of training, time since training, or when training was done. The last of these findings implies that there was little improvement over time in the operation of training programs for men.

Finally, the estimates for men indicate that there is a substantial random component in the estimated effects of training, but the covariates reduce this random component substantially. In Model 1, over one-half the estimated error variance is due to the random component. In Models 5 through 8, in contrast, less than 20% of the error variance is due to the random component.

Women

The results for women, which are reported in Table 6, show clear differences across the various training types that are very robust with respect to model specification. Three of these training types—classroom skills training, OJT, and mixed classroom and workplace training—are associated with increases in earnings that are well above $1,000 per year, while subsidized work results in somewhat smaller, but still substantial, increases. Training involving basic education, on the other hand, appears to have been ineffective. However, only 7 of the 133 observations for women were for this type of training. The low p-level in Model 1 indicates that these differences among training types are statistically significant.

Few of the other covariates appear to be significantly related to the training effect for women. For example, labor market conditions, as measured by the unemployment rate and the percentage of the work force in manufacturing, appear to have had little influence on the extent to which training increased earnings. The coefficients on these variables are usually small and are never statistically significant. We also examined whether the maximum AFDC payment for a family of three 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) was positive but relatively small and never approached conventional levels of statistical significance.[21] Similarly, there is no evidence that more expensive training programs performed better than less expensive ones.

The lack of a negative coefficient on the time since training variable suggests that for women, as for men, the effects of training persisted for at least several years after the training was complete. In fact, there is weak evidence that the effects increased, but this effect is statistically significant only in Model 6. In contrast to the results for men, there is evidence that earlier training programs had larger effects than more recent programs. Also in contrast to the findings for men, there is no evidence that experimental and nonexperimental evaluations produced systematically different results, suggesting that any biases in nonexperimental estimates tend to cancel

[21]Because the AFDC payment level variable was not statistically significant for women and was not used in the regressions for men and youths, we excluded it from the regression specifications reported in Table 6 to facilitate comparisons with the results in Tables 5 and 7.

Table 6. Variation in Program Effects for Women: Mixed Effects Model Results. (Standard Errors in Parentheses)

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Training Type								
Classroom Skills Training	1,787***	1,702***	1,672***	1,594***	1,518***	1,306***	1,295***	1,335***
	(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	1,619***	1,568***	1,594***	1,591***	1,637***	1,443***	1,570***	1,644***
	(241)	(240)	(238)	(242)	(243)	(252)	(253)	(325)
Subsidized Work	816***	880***	1,010***	1,088***	1,051***	811***	848***	784***
	(228)	(227)	(234)	(242)	(242)	(260)	(255)	(269)
Mix of Classroom and Workplace Training	1,410***	1,430***	1,436***	1,490***	1,479***	1,331***	1,580***	1,538***
	(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 ($1-[v^*/\sigma^2]$)	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 errors 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%. *** Significant at 1% level.

out across studies.

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% of the total error variance). For women, in other words, a large part of the variation in the effects of training cannot be explained by either sampling error or observed variables. As was shown earlier, however, there is less variation in the overall distribution of training effects for women than for men or youths, so in a sense there is less to explain.

Youths

We noted earlier that the overall effect of training for youths was close to zero. Table 7 suggests, however, that certain types of training may nonetheless have been effective for youths. Specifically, classroom skills training seems to have had a positive payoff, with estimated effects exceeding $1,400 in each specification. While the coefficients on OJT and subsidized work are also positive, they are fairly small and always statistically insignificant.

Perhaps because there is considerably more variation in the estimated training effects for youths than for men and women, several other covariates achieve statistical significance as factors explaining this variation. For example, training appears to have been less effective for whites than for nonwhites or for the omitted mixed groups of whites and nonwhites. 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, though they are not statistically significant, suggesting that the payoff from training was larger for male youths (the omitted group).

Training for youths might be ineffective because youth unemployment is usually very high. The findings in Table 7 suggest that training effects for youths were highly sensitive to the unemployment rate. In particular, training appears to have become less effective as unemployment increased. Although the quadratic terms indicate the effects of training would begin increasing if unemployment increased enough, this upswing does not occur until unemployment reaches 45%, which is the maximum unemployment rate in the sample. In Model 4, for example, at the mean unemployment rate in the sample (22.3%), 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%, the model predicts the effect would be only $86.

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

Finally, the random component for youths is proportionately smaller than that for men and much smaller than that for women. As in the results for men (but not for women), it declines in importance as additional covariates are added to the model. In Model 1, the random component comprises about 25% of the total error variance, while in Model 8, it comprises only about 6% of the total error variance.

Sensitivity Tests

We performed two tests to determine whether the results reported in this paper were sensitive to the assumptions made in estimation. First, because the standard error of the estimated training program effect plays such a crucial role in the meta-analysis, it is important to determine whether the results are sensitive to the method used to impute standard errors in cases where they are missing from the evaluation study. To investigate this issue, we re-estimated Model 8 using two alternatives to the procedure used in Tables 5–7. In the first re-estimation, we simply excluded the 71 cases in which the standard errors were not provided in the original study. This reduced the sample size and hence the precision of the estimated coefficients in Model 8.

In the second re-estimation, we imputed missing standard errors randomly using the following procedure. First, we ran a regression of the logarithm of the standard error on the absolute value of the training program effect for the cases in which the stan-

Table 7. Variation in Program Effects for Youths, Mixed Effects Model Results. (Standard Errors in Parentheses)

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Training Type								
Classroom Skills Training	1,547*** (445)	1,524*** (455)	1,430*** (387)	1,581*** (355)	1,563*** (360)	1,634*** (362)	1,599*** (382)	1,954*** (377)
CT+Basic Ed	–377 (834)	–431 (862)	–537 (638)	–124 (521)	–141 (521)	–86 (518)	–115 (540)	–1,006* (498)
OJT	277 (584)	251 (593)	172 (545)	344 (524)	332 (527)	405 (529)	366 (543)	709 (539)
Subsidized Work	75 (305)	58 (313)	–60 (254)	69 (231)	48 (344)	319 (392)	423 (446)	273 (417)
Mix of Classroom and Workplace Training	–323 (199)	–327 (200)	–335* (167)	–206 (152)	–207 (151)	–26 (195)	–19 (198)	–91 (186)
Female	69 (250)	77 (253)	92 (210)	–202 (203)	–194 (219)	–237 (220)	–261 (227)	–247 (212)
Men and Women	201 (391)	217 (398)	173 (321)	–45 (293)	–29 (313)	–77 (313)	–225 (448)	–658 (425)
Experimental Dummy		68 (243)	29 (243)	–40 (221)	–10 (288)	–48 (288)	–197 (413)	186 (412)
White			–1,070*** (305)	–1,147*** (284)	–1,136*** (313)	–1,017*** (322)	–993*** (335)	–1,279*** (321)
Non-White			470 (288)	–239 (478)	–221 (497)	–404 (512)	–396 (520)	–458 (527)
Unemployment Rate				–232*** (67)	–230*** (68)	–219*** (68)	–229*** (72)	–255*** (68)
Unemployment Rate Squared				5*** (1)	5*** (1)	5*** (1)	5*** (1)	5*** (1)
Percent Manufacturing Employment					2 (40)	–8 (40)	0 (46)	–78* (47)
Years Since Training						–140 (96)	–130 (101)	–116 (92)
Year of Training							16 (37)	–13 (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 ($1–[v^*/\sigma^2]$)	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 errors in parentheses. All variables except training types and years since training are centereed 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.

dard errors were not missing. Separate regressions were run for each of the four significance levels (not statistically significant and statistically significant at the 10%, 5%, and 1% levels, respectively). Then the regressions were used to predict the log of the standard error for program effect estimates for cases with missing standard errors, adding a random term with a standard deviation equal to the standard error of the regression. The resulting estimate was then exponentiated to obtain a predicted standard error. A table that compares findings from the two alternatives to those from the original estimates is available upon request from the authors.

With one exception, the results were quite robust with respect to the alternative imputation methods. The exception was for men with missing cases excluded from the regression sample. In this case, several (but not all) of the coefficients changed substantially. When the missing cases were imputed randomly for men, however, the results were very close to the original estimates. For women and youths, all three methods of imputation yielded similar results.

We also tested the sensitivity of our results to extreme values by re-estimating our models excluding the highest and lowest three estimated program effects. For the most part, the results were substantively unchanged. Although the magnitudes of the coefficients changed somewhat, their signs and levels of statistical significance were generally similar. However, a few of the coefficients changed considerably, particularly for youths. Again, a table with these results is available from the authors upon request.

Conclusions

We have presented findings from a meta-analysis of voluntary government training programs for the disadvantaged. Exploiting the fact that there are numerous estimates of program effects on earnings, we examined the influence of a number of factors on the effectiveness of training.

Men, women, and youths were studied separately. The factors explaining variation in program effects 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. Our findings are most robust for women, the group for whom we have the most training effects estimates, and least robust for men, the group for whom we have the fewest estimates.

Although the United States has over three decades of experience in running training programs, our meta-analysis suggests that these programs have not become more effective over time. Moreover, the effects are rarely found to be large. Even for adult women, the vast majority of estimates indicate that training programs increased earnings by less than $2,000 a year for a typical trainee. However, compared to the average cost of about $6,600, effects close to $2,000 may be sizable if they persist for several years, and our findings suggest that they did.

The findings suggest that more expensive training programs were not necessarily superior ones for adults. This is a surprising result and suggests either that funds were not being well spent or that factors influencing the effectiveness of programs were also influencing their costs. For example, more expensive programs may have dealt with more disadvantaged trainees.

Among types of training, none were found to be consistently superior. Basic education was generally found to be ineffective, however, while classroom skills training was almost always found to be effective. For women, most types of training (with the possible exception of basic education) seem to have been effective. In contrast, only classroom skills training appears to have been effective for youths, and the findings on effectiveness of training types for men were not robust enough to draw conclusions.

We found no evidence that a higher unemployment rate made training more effective, except possibly at very high levels of unemployment. There is, however, some evidence supporting 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 to have been less effective for white men and white youths than for nonwhite and racially mixed groups of men and youths. One possible explanation for this finding, which did not occur for women, is that white workers faced fewer employment barriers than nonwhite workers did and could more readily find jobs on their own without the aid of training.

Given the findings summarized above, it may be useful to ask what was learned from this meta-analysis that was not previously known. Previous syntheses of the training literature (LaLonde 1995; Friedlander, Greenberg, and Robins 1997; Heckman, LaLonde, and Smith 1999) have documented that voluntary government-funded training programs rarely produce large effects on earnings, that these effects are larger for adult women than for adult men and are negligible for youths, and that within each of these three groups estimates of the effects vary considerably. Moreover, according to Friedlander, Greenberg, and Robins (1997), "the link between increased cost … and greater earnings effect has not been firmly established" (p. 1834), "the limited evidence available suggests that earnings effects may persist" after training is completed (p. 1836), and some evidence exists that classroom skills training may be more effective than basic education (p. 1836). Although these conclusions are all consistent with our findings, they were not formally tested in earlier syntheses. For example, Friedlander, Greenberg, and Robins's conclusion about the superiority of classroom skills training over basic education was based on findings from only two evaluations (the Minority Female Single Parent demonstration and the Jobstart demonstration) in which a single site that emphasized skills training for specific occupations produced larger earnings effects than the other evaluation sites, all of which focused more on basic education.

There are several other ways in which this study covers new ground. Standard surveys of the training program evaluation literature do not seem to have drawn conclusions about whether training programs have become more effective over time. In addition, except for differences between men, women, and youths, little seems to be known about the types of trainees for whom training programs work best. Moreover, there has been virtually no systematic attempt to study how macro-economic conditions affect program success. As indicated by this study, such issues can be fruitfully examined through meta-analysis. For example, we found evidence that training programs are less successful for some groups at higher levels of unemployment. Two recent studies of mandatory welfare-to-work programs that also used meta-analytic tools (Bloom, Hill, Riccio 2001; Ashworth et al. Forthcoming) similarly found that these programs produce smaller earnings effects when the unemployment rate is higher.

REFERENCES

I. Studies Used in the Meta-Analysis

Ashenfelter, Orley. 1978. "Estimating the Effects of Training Programs on Earnings." *Review of Economics and Statistics,* Vol. 60, No. 1 (February), pp. 47–57.

Ashenfelter, Orley, and David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs." *Review of Economics and Statistics,* Vol. 67, No. 4 (November), pp. 648–60.

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

Bassi, Laurie J. 1983. "The Effect of CETA on the Postprogram Earnings of Participants." *Journal of Human Resources,* Vol. 18, No. 4 (Fall), pp. 539–56.

____. 1984. "Estimating the Effect of Training Programs with Non-Random Selection." *Review of Economics and Statistics,* Vol. 66, No. 1 (February), pp. 36–43.

Bell, Stephen H., and Larry L. Orr. 1984. "Is Subsidized Employment Cost Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations." *Journal of Human Resources,* Vol. 19, No. 1 (Spring), pp. 42–61.

Bloom, Howard S. 1984. "Estimating the Effect of Job Training Programs Using Longitudinal Data: Ashenfelter's Findings Reconsidered." *Journal of Human Resources,* Vol. 19, No. 4 (Fall), pp. 544–56.

____. 1987. "What Works for Whom? CETA Impacts for Adult Participants." *Evaluation Review,* Vol. 11, No. 4 (August), pp. 510–27.

Bryant, Edward C., and Kalman Rupp. 1987. "Evaluating the Impact of CETA on Participant Earnings." *Evaluation Review,* Vol. 11, No. 4 (August), pp. 473–92.

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

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

Cave, George et al. 1993. *JOBSTART: Final Report on a Program for School Dropouts.* New York: MDRC.

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

Couch, Kenneth A. 1992. "New Evidence on the Long-Term Effects of Employment and Training Programs." *Journal of Labor Economics,* Vol. 10, No. 4 (October), pp. 380–88.

Dickinson, Katherine P., Terry R. Johnson, and Richard W. West. 1984. *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, Calif.: SRI International.

____. 1986. "An Analysis of the Impact of CETA Programs on Participants' Earnings." *Journal of Human Resources,* Vol. 21, No. 1 (Winter), pp. 64–91.

____. 1987a. "The Impact of CETA Programs on Components of Participants' Earnings." *Industrial and Labor Relations Review,* Vol. 40, No. 3 (April), pp. 430–40.

____. 1987b. "An Analysis of the Sensitivity of Quasi-Experimental Net Impact Estimates of CETA Programs." *Evaluation Review,* Vol. 11, No. 4 (August), pp. 452–72.

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

Freedman, Stephen, Jan Bryant, and George Cave. 1988. *New Jersey: Final Report on the Grant Diversion Project.* New York: MDRC.

Gay, Robert, and Michael Borus. 1980. "Validating Performance Indicators for Employment and Training Programs." *Journal of Human Resources,* Vol. 15, No. 1 (Winter), pp. 29–48.

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

Kiefer, Nicholas M. 1978. "Federally Subsidized Occupational Training and the Employment and Earnings of Male Trainees." *Journal of Econometrics,* Vol. 8, No. 1, pp. 111–25.

____. 1979. "The Economic Benefits from Four Government Training Programs." In Farrell E. Bloch, ed., *Research in Labor Economics* (Supplement 1), pp. 159–86.

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

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

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

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

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

Westat. 1984. *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.

Zambrowski, Amy, and Anne Gordon. 1993. *Evalua-*

tion of the Minority Female Single Parent Demonstration: Fifth-Year Impacts at CET. Princeton: Mathematica Policy Research.

II. Other References

Ashworth, Karl, Andreas Cebulla, David Greenberg, and Robert Walker. Forthcoming. "Meta-Evaluation: Discovering What Works Best in Welfare Provision." *Evaluation.*

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

Cohen, Jacob. 1988. *Statistical Power Analysis in the Social Sciences.* Hillsdale, N.J.: Erlbaum.

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

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

Fraker, Thomas, and Rebecca Maynard. 1987. "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs." *Journal of Human Resources,* Vol. 22, No. 2 (Spring), pp. 194–227.

Friedlander, Daniel, David H. Greenberg, and Philip K. Robins. 1997. "Evaluating Government Programs for the Economically Disadvantaged." *Journal of Economic Literature,* Vol. 35, No. 4 (December), pp. 1809–55.

Friedlander, Daniel, and Philip K. Robins. 1995. "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods." *American Economic Review,* Vol. 85, No. 4 (September), pp. 923–37.

Greenberg, David, Charles Michalopoulos, and Philip K. Robins. Forthcoming. "What Happens to the Effects of Government-Funded Training Programs over Time?" *Journal of Human Resources.*

Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. "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, pp. 1865–2097.

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

____. 1992. "Meta-Analysis." *Journal of Educational Statistics,* Vol. 17, No. 4 (Winter), pp. 279–96.

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

Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis.* Boston: Academic Press.

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

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

Jarrell, S. B., and T. D. Stanley. 1990. "A Meta-Analysis of the Union-Nonunion Wage Gap." *Industrial and Labor Relations Review,* Vol. 40, No. 2 (January), pp. 54–67.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review,* Vol. 76, No. 4 (September), pp. 604–20.

____. 1995. "The Promise of Public Sector-Sponsored Training Programs." *Journal of Economic Perspectives,* Vol. 9, No. 2 (Spring), pp. 149–68.

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

Moffitt, Robert. 1991. "Program Evaluation with Nonexperimental Data." *Evaluation Review,* Vol. 15, No. 3 (June), pp. 291–314.

Raudenbush, Stephen W. 1994. "Random Effects Models." In Harris N. Cooper and Larry V. Hedges, eds., *The Handbook of Research Synthesis.* New York: Russell Sage Foundation, pp. 301–21.

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