The Effect of Direct Job Creation and Training Programs on Low-Skilled Workers

Laurie J. Bassi
Georgetown University

and

Orley Ashenfelter
Princeton University

July 1985

This research was supported in part by a contract to the Institute for Research on Poverty of the University of Wisconsin from the U.S. Department of Health and Human Services. The views expressed in this paper are those of the authors and do not necessarily reflect the views of either DHHS or of the Institute for Research on Poverty. Joseph Antos and Peter Gottschalk provided helpful comments on an earlier draft. It is scheduled to appear as a chapter of Sheldon H. Danziger and Daniel H. Weinberg, eds. "Alleviating Poverty: What Works and What Does Not."
Abstract

This paper summarizes the literature on the impact of employment and training programs and concludes that they have been neither an overwhelming success nor a complete failure in terms of their ability to increase the long-term employment and earnings of disadvantaged workers. Employment and earnings programs capacity to improve the lot of any given participant, and the collective economic well-being of the disadvantaged, has been modest — as has been the level of resources devoted to these programs.

Two decades of non-experimental program evaluation in the employment and training field have finally taught a lesson about which there can be little disagreement: Convincing program evaluation is going to require continued use of randomized trials. We wish to emphasize that this is not simply a statement that randomization is a preferable methodological approach regardless of the field of study. Instead, we believe the evidence in the study of employment and training programs overwhelmingly indicates that randomization is essential in program evaluation in the employment and training field. The difficulty seems to be that the earnings determination processes of today's workers and the program selection methods of today's programs interact to make it nearly impossible to produce reliable estimates of how workers earnings would have behaved in the absence of a program. The evidence to support this finding comes from both experimental and non-experimental studies. The non-experimental studies indicate that minor changes in methods, for which there is no empirical justification, produce large swings in estimated program effects. The study of experimental findings indicates that perfectly plausible non-experimental methods may lead to dramatic errors in inferences about program effects.

This basic finding raises a fundamental question: What is the proper reaction of policy makers? In our view the appropriate reaction is for policymakers to begin the development of a credible research and development effort using randomized clinical trials in a wide variety of study areas.
For more than two decades now, the federal government has been continuously involved in some type of active intervention in the labor market. The debate over the appropriate size and method of intervention has been long, heated, emotional, and political. In some instances federally funded employment and training programs have clearly been used to appease special interest groups; in others, because of their widespread political popularity. The central economic rationale for federally funded employment and training programs has often been lost against this political backdrop.

From an economic perspective, there are only three possible reasons for the government to intervene in the operation of the labor market. One rationale is to reduce frictional unemployment. The Employment Service has long been in business for exactly this purpose. The second rationale is to reduce cyclical unemployment. We would expect that programs that are countercyclical in nature would be turned on and off in response to changes in the unemployment rate. The final rationale is to alleviate structural unemployment -- that unemployment which is in some sense "involuntary" and persists over the course of the business cycle. Programs that are structural in nature should have targeting mechanisms to assure that the individuals served are persistently unemployed or underemployed.

This paper concentrates on what is known about the effectiveness of programs that are primarily structural in nature. As will become evident, however, it is not always possible to clearly distinguish between countercyclical and structural programs. We begin with a review of the historical development of employment and training policy in the United States in Section I, and examine how funding levels for these programs have changed over time. In Section II we review the evidence on what is known about the effectiveness of these programs and consider the costs of alternative types of programs. In
Section II we consider the implications that the size of the programs as well as fiscal substitution rates, have for their aggregate impact. In Section III we present our summary and conclusions.

I. A History of the Role of Government Employment and Training Programs in the U.S. Labor Market

The history of federal involvement in employment and training programs can be divided into five phases. The first phase evolved out of the unemployment crisis of the Great Depression. The Works Progress Administration (WPA), created in 1935, was a massive direct job creation program. During 1936 the program’s peak year, WPA employed over three million out of the total nine million who were unemployed (Howard, 1943). At the same time the Public Works Administration (PWA) concentrated on capital-intensive projects in the public sector on the assumption that this spending would generate increased private sector activity. These programs were phased out by 1943.

After this initial involvement in the labor market, there was a long period of no intervention. Only in the early 1960s did public employment programs once again receive widespread attention. This second phase had a much different emphasis (and life-span) than did the first. It grew out of a concern for workers who had been displaced by technological advances as well as a concern over employment bottlenecks that these advances generated. The Manpower Development and Training Act (MDTA), passed in 1962, was originally designed to provide vocational and on-the-job training for displaced workers. Initially the program primarily served male heads of households with
substantial previous labor market experience. However, the emphasis of the program quickly changed to meet the needs of more disadvantaged individuals, those with "hard-core" unemployment problems.

This shift in emphasis ushered in the third phase of federal involvement, coinciding with the implementation of Great Society programs and lasting through the early 1970s. Employment and training programs were increasingly targeted at minorities, welfare recipients, low-income youth, the elderly, and other hard-to-employ groups. The plethora of programs included: Job Corps, Neighborhood Youth Corps, Operation Mainstream, New Careers, Concentrated Employment Program, Older Americans, Model Cities, Foster Grandparents, and the Work Incentive Program. These programs provided work experience and training (both on-the-job and classroom), with the intention of improving the long-term employability of participants and providing career ladders for moving from temporary public sector jobs to permanent jobs in the private sector. The emphasis was structural in nature.

With the recession of 1970-71, public attention began to shift from the long-term employability problems of the disadvantaged to the problems of the cyclically unemployed. Under the Emergency Employment Act (EEA) of 1971, the federal government provided $1 billion during fiscal year 1972, and $1.25 billion during 1973 for the purpose of direct job creation within state and local governments under the Public Employment Program (PEP). These funds were allocated in proportion to the number of unemployed within each area. The disadvantaged, however, were not completely forgotten. Provisions were made for targeting a variety of groups including: Vietnam veterans, youth, elderly, migrants, non-English speaking persons, and welfare recipients. In addition, a limited amount of money was made available for training. The passage of the Emergency Employment Act represents the beginning of the fourth phase of
federal involvement in employment and training policy. This phase, which
continued through 1978, was characterized by a mixed strategy that attempted
to combat both cyclical and structural unemployment.

This strategy continued with the passage of the Comprehensive Employment
and Training Act (CETA) of 1973, which consolidated many of the training
programs of the late 1960s and early 1970s, and also incorporated the Public
Employment Program. When CETA passed, unemployment was well below its 1971
peak, and its original emphasis was clearly on training rather than
employment. However, some provisions for public service employment (PSE) were
made for high unemployment areas. The deep recession of 1974-75 resulted in a
change of priorities under CETA, with countercyclical PSE occupying center
stage. Very substantial countercyclical expenditures continued to be made
through 1978 when CETA was reauthorized.

During this period employment and training programs were very well-funded
by historic standards, and considerable experimentation was being
undertaken. It was during this time that a wide variety of alternative
policies were used on a trial basis. The first of these was the Supported
Work Demonstration, one of the few employment/training programs that has been
run as an experiment with a randomly selected control group. The
demonstration tested the effects of a highly structured work experience on
four disadvantaged target groups: long-term AFDC recipients, ex-addicts, ex-
offenders, and young school dropouts. Supported Work was distinguished from
other programs by its emphasis on gradually bringing individuals with extreme
employment disabilities into the "world of work" by using peer group support,
graduated stress, and close supervision as program techniques.

The second major development in this period was the Youth Employment and
Demonstration Act (YEDPA) passed in 1977, which funded a variety of programs
specifically directed to the needs of youth. These programs were administered under the CETA umbrella, and provided large budgets for research and evaluation.

A third development was the use of the targeted Jobs Tax Credit to subsidize (and hopefully increase) employment among certain disadvantaged groups.

The 1978 reauthorization of CETA represents the entrance of the phase of employment and training policy in which we currently find ourselves. During the CETA reauthorization, funding for countercyclical PSE was cut drastically, partly in response to a dramatic decline in unemployment and partly in response to widespread allegations of abuse and fraud under the PSE program. CETA once again became almost exclusively structural in nature.

In late 1982, CETA was replaced by the Jobs Training Partnership Act (JTPA) which emphasizes combatting structural rather than cyclical unemployment. No funds were made available for any form of direct job creation despite the fact that the unemployment rate had reached double-digit levels while the JTPA legislation was being created. The Reagan administration opposes direct job creation in the public sector, believing strongly that the only federal responsibility is to provide training for disadvantaged individuals. This philosophy is reflected in the JTPA legislation.

We have now come full circle and returned to the types of programs we had in the mid-1960s. Our programs are structural in nature, providing training for the most disadvantaged members of the labor force. However, over the course of these two decades, much has changed. Table 1 shows real per capita funding levels for counter-cyclical and structural programs and unemployment rates over the past two decades. These data indicate that funding levels for
structural programs increased steadily from the early 1960s through 1979, when they dropped sharply, and funding levels for countercyclical programs have gyrated wildly through the 1970s. This table also indicates that changes in countercyclical funding have not been well timed in terms of changes in the unemployment rate.

The empirical relationships between funding levels and unemployment rates were examined using regression analysis.\textsuperscript{1} An interesting result that emerges from this analysis is that funding levels for both structural and countercyclical programs were higher when the Democrats were in office. The analysis also confirms that funding levels have, indeed, been very poorly coordinated with the unemployment rate.

While this finding is not too surprising for structural funding levels, since these programs are intended to combat structural rather than cyclical unemployment, it does indicate that the funding levels for countercyclical programs have not been very sensible from an economic perspective. In order to be effective from a macroeconomic perspective, it would be necessary for funding levels to be highly correlated with current unemployment. However, it is lagged unemployment, rather than current unemployment, that is a much better predictor of countercyclical funding.
Table 1
Real, Per Capita Funding Levels for Employment and Training Programs

<table>
<thead>
<tr>
<th>Year</th>
<th>Unemployment Rate</th>
<th>Real, Per Capita Structural Funding ($)</th>
<th>Real, Per Capita Counter-Cyclical Funding ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1963</td>
<td>5.48</td>
<td>.004</td>
<td>0</td>
</tr>
<tr>
<td>1964</td>
<td>5.31</td>
<td>.010</td>
<td>0</td>
</tr>
<tr>
<td>1965</td>
<td>4.64</td>
<td>.029</td>
<td>0</td>
</tr>
<tr>
<td>1966</td>
<td>3.84</td>
<td>.042</td>
<td>0</td>
</tr>
<tr>
<td>1967</td>
<td>3.68</td>
<td>.051</td>
<td>0</td>
</tr>
<tr>
<td>1968</td>
<td>3.61</td>
<td>.048</td>
<td>0</td>
</tr>
<tr>
<td>1969</td>
<td>3.33</td>
<td>.059</td>
<td>0</td>
</tr>
<tr>
<td>1970</td>
<td>4.17</td>
<td>.076</td>
<td>0</td>
</tr>
<tr>
<td>1971</td>
<td>5.72</td>
<td>.075</td>
<td>0</td>
</tr>
<tr>
<td>1972</td>
<td>5.67</td>
<td>.083</td>
<td>.046</td>
</tr>
<tr>
<td>1973</td>
<td>4.96</td>
<td>.068</td>
<td>.055</td>
</tr>
<tr>
<td>1974</td>
<td>4.94</td>
<td>.076</td>
<td>.011</td>
</tr>
<tr>
<td>1975</td>
<td>7.69</td>
<td>.095</td>
<td>.057</td>
</tr>
<tr>
<td>1976</td>
<td>7.75</td>
<td>.097</td>
<td>.079</td>
</tr>
<tr>
<td>1977</td>
<td>7.28</td>
<td>.109</td>
<td>.200</td>
</tr>
<tr>
<td>1978</td>
<td>6.23</td>
<td>.154</td>
<td>.066</td>
</tr>
<tr>
<td>1979</td>
<td>5.74</td>
<td>.199</td>
<td>.090</td>
</tr>
<tr>
<td>1980</td>
<td>6.57</td>
<td>.175</td>
<td>.041</td>
</tr>
<tr>
<td>1981</td>
<td>7.27</td>
<td>.137</td>
<td>.029</td>
</tr>
<tr>
<td>1982</td>
<td>8.77</td>
<td>.095</td>
<td>.002</td>
</tr>
<tr>
<td>1983</td>
<td>9.25</td>
<td>.089</td>
<td>0</td>
</tr>
<tr>
<td>1984</td>
<td>7.86</td>
<td>.064</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Funding levels are for Department of Labor financed programs and are expressed in constant 1972 dollars. The sources of these data are as follows: unemployment rates and the GNP price deflator were taken from the Economic Report of the President, and funding levels were from various issues of the Employment and Training Report. Funding levels for 1982-1984 are estimates provided by Demetra Nightingale of the Urban Institute. These figures represent federal obligations for all work and training programs administered by the Department of Labor. Obligations for the Public Employment Program and the Public Service Employment Program were defined as being countercyclical. All other obligations were considered to be structural.
Critics of countercyclical programs have long pointed to this phenomenon. The political system is unable to respond quickly enough for the programs to be timely. It could, in fact, be the case that such programs are actually counterproductive — turning on the faucet just when it needs to be turned off.

The poor timing of countercyclical programs was, no doubt, at least partially responsible for their ultimate demise. By the time the programs were in full swing, our concerns had usually switched from reducing countercyclical unemployment to combating structural unemployment. Inevitably, countercyclical programs have been judged by an evaluation criterion that would be more appropriate for judging the effectiveness of structural programs.

With this discussion by way of background, we turn now to an examination of that criterion — the extent to which employment and training programs have increased the earnings' capacity of their participants. We remind the reader of our primary emphasis on structural programs, even though, it may not be valid to judge countercyclical programs within this context.

II. The Effect of Employment and Training Policies on the Employment and Earnings of Low Wage Workers

During the past two decades, billions of dollars have been spent on employment and training programs, and millions have been spent on research, evaluation, and demonstrations. It would be a heroic task to summarize and integrate the results of all this work, and one which would probably not be
tremendously fruitful. In this section we concentrate on some of the major studies of the most fundamental aspects in the evaluation of the effectiveness of employment and training programs. The first is the estimation of the microeconomic impact of the programs. In particular, what have been the employment and earnings impacts on participants. The second is the determination of the macroeconomic effect of the programs, i.e., their net employment effect. We review both sets of studies in turn.

The Microeconomic Effects of Employment and Training Programs

Much, if not most, of the analytic work in this area is woefully inadequate. The reasons for this are many. Perhaps the most important is that almost no employment or training programs have been run on an experimental basis. Consequently, researchers have been forced either to create artificial "comparison groups" or to conduct their analysis without one. In the latter case, the impact of the program on participants' post-program employment and earnings experience is simply estimated by comparing their pre- and post-program experience, with no regard for the effect of the passage of time, or changing economic conditions.

Another problem which has plagued evaluations of employment and training programs is what has come to be known as the "pre-program dip." It is generally the case that, immediately prior to program entry, participants experience a marked decline in employment and earnings. This is no doubt in large part because many of the programs have required, as a condition of eligibility, that participants be unemployed. Since it is impossible to know whether this dip in earnings is merely transitory or the beginning of a permanent decline, it is difficult to know what a participant's permanent
earnings level was prior to the program. And it is not easy to find comparison group members that experience the same decline in employment and earnings.

Even if it is possible to generate a comparison group that is comparable to the participants based on observable characteristics that affect earnings and employment, it is very likely that the comparison group members differ from the participants in a systematic but unobservable fashion. Participants, for instance, may be systematically more (or less) "motivated" than their comparison group counterparts. Since "motivation" is not observable, and it is likely to be an important determinant of earnings, it is unclear the extent to which post-program earnings differentials between participants and comparisons results from program participation or differences in motivation. Consequently, in the absence of an experimental design that randomly assigns individuals to either a treatment or control group, it is necessary to employ statistical techniques in an attempt to control for nonrandom assignment due to unobservable characteristics.

Finally, researchers have been confronted with "contamination" of the comparison groups that are available. Contamination occurs when members of the comparison groups are, or at some point have been, program participants. In general, there is no way to identify which members of the comparison groups have been "contaminated." This introduces an errors in variable problem since program participation is measured inaccurately for contaminated individuals.²

Very few evaluations of employment and training programs have had well-chosen comparison groups, adjusted for contamination, and simultaneously controlled for nonrandom selection. In the review of the evaluation literature that follows, we confine our discussion to a few of the better evaluations, even though most of these have serious shortcomings. But after
presenting our review, we summarize what conclusions can be drawn. The review will proceed on a chronological basis.

Ashenfelter (1978) did one of the first evaluations of the MDTA program that attempts to meet most of the criteria specified above. Using a comparison group drawn from the Continuous Work History Sample, he estimated the effect of participation in classroom training for all participants who entered the MDTA program during the first three months of 1964. The outcome measure of program participation was annual Social Security earnings in the first five post-program years. Ashenfelter found that MDTA classroom training did, indeed, have a positive and statistically significant effect on participants’ earnings ranging from $200 per annum for white males up to about $550 for black females. In general, the effects were larger for women than for men — a finding that appears consistently. Ashenfelter also found some evidence of a decay in the earnings effect over time for men, but not for women. He points out, however, that the decay effect (particularly for men) is sensitive to the estimation procedure that is used. This sensitivity points to the central dilemma that confronts researchers. In the absence of experimental data, the estimated benefits of programs, and the pattern of these benefits over time, may vary dramatically depending on the estimation strategy that is chosen. And the decay rate of these effects, or the lack thereof, has important implications for the cost-effectiveness of training programs.

Using different cohorts of MDTA participants and alternative comparison groups drawn from other samples, two other studies have found widely differing earnings gains for men. Kiefer (1978) found large negative effects while Cooley, McGuire, and Prescott (1979) found large positive effects. The estimated earnings gains for women, however, are much less sensitive to the
choice of the cohort and the comparison group.

After the MDTA program, the next major employment/training program was the Public Employment Program, which was primarily countercyclical but also had some structural objectives. Because it is in some sense inappropriate to judge countercyclical programs, which have income maintenance goals, by criteria that are intended for structural programs, we should not be too surprised if they do not generate long-term employment and earnings gains.

Westat (1979) has done the only major analysis of the net earnings impact of the Public Employment Program (PEP). The Current Population Survey (CPS) was used with matched Social Security earnings files to generate comparison groups for PEP participants. Using both an autoregressive earnings model and a comparison of mean earnings between participants and comparisons, Westat came to similar conclusions as did Ashenfelter. While in-program earnings gains were substantial for all race/sex groups, women generally experienced greater post-program gains than did men; the gains were also larger for minorities than whites. The estimated earnings gains for men were not significantly different from zero.

The availability of the CPS with matched Social Security files represented a methodological advance that allows researchers to inexpensively generate a variety of comparison groups. In addition, the recently available Continuous Longitudinal Manpower Survey (CLMS) represents a major data development effort of the Department of Labor for evaluation of CETA-financed employment and training activities. The public-use tapes contain large, representative samples of individuals who participated in the program during fiscal years 1975 through 1979. For each of these individuals, the CLMS included a four-year record of labor force experience beginning one year prior to CETA enrollment, the type of CETA program(s) in which the individual
participated, basic demographic characteristics, a history of public benefits received by the individual and/or the individual's family, and family-related variables. Although the CLMS is the best available data base, it is still less than ideal -- for example, variables that are often included in earnings equations, such as union status, geographic location, and work experience, are not included. However, the unique advantage of having both the CLMS and the CPS available with merged records from the Social Security Administration is that researchers can now generate a wide array of comparison groups, facilitating testing of the sensitivity of estimated program effects to the choice of the comparison group. And because longitudinal data are available, researchers have used autoregressive or fixed effects models to control for nonrandom selection.

Not surprisingly, this expanded ability to experiment with comparison groups and estimation techniques has led to a broad array of estimated program effects. Some consistent findings do emerge, however, from this growing literature. As was the case with evaluations of earlier programs, CETA evaluations have found with remarkable consistency that women benefit more from program participation than do men. In fact, CETA appears to have generated no significant earnings gains for men whatsoever. It seems that participation in employment and training programs increases earnings primarily through an increase in hours worked rather than through an increase in wages (Bassi et al. 1984, and Levy 1982). Since women generally work fewer hours than men, there is obviously more room for an impact on their hours of work than is the case for men.

Related findings indicate that the individuals who benefited the most from CETA participation were among the most disadvantaged with the least amount of previous labor market experience. This suggests that CETA has been
successful in preparing participants for entry level positions. While this may represent a substantial improvement for many female participants, it may leave many male participants no better off. Consistent with this hypothesis is Fraker, Maynard, and Nelson's (1984) finding that young men are the only subgroup of men who do experience substantial post program employment and earnings gains.  

A final finding from the CETA evaluation literature is that the least effective method of increasing participants post-program employment and earnings is through "work experience" programs. There were four basic types of programs available under the auspices of CETA: classroom training, on-the-job training, public service employment, and work experience. The work experience program seemed to be reserved for the most disadvantaged participants. Apparently, this activity involved a good deal of acclimation to the "world of work" and not much in the way of training or working at a "real job." This was very different from the public service employment program where participants often worked side by side with regular civil service employees. While this difference may account for the differential post-program effects found under these two alternative types of employment programs, it is also possible that the difference merely reflects a selection bias.

The other three program activities: classroom training, on-the-job training, and public service employment all resulted in positive and significant post-program employment and earnings gains for women. There is, however, no clear agreement on which program activity was the most effective. In general earnings gains for women in a two to three year period after the program are in the range of $600-$1200, with most estimates in the upper end of this range. The gains show no sign of decay in the post-program
period, although few years of post-program data are available. Since most researchers have not adjusted for contamination bias, these estimates are best considered as lower bounds and as very substantial in light of the average earnings levels of female participants (about $2700 in the pre-program period).

None of the programs have a consistently significant effect on men. Occasionally, on-the-job training positively affects some subgroups of men. But because there appears to be significant "creaming" of participants into on-the-job training, the validity of this result is in serious question.

There is also an expanding research on programs addressing the employability problems of youth. Perhaps the best known, and certainly the most enduring of these programs, is the Job Corps. Job Corps, in continuous operation since 1965, serves limited numbers of the most disadvantaged youth, ages 16 to 21. The two types of Job Corps centers, residential and non-residential, both involve very intensive (and expensive) evaluations of each participant's barriers to employment and provide extensive services. These include individual and group counseling, medical attention, remedial education, vocational training, GED courses, and courses in citizenship.

There have been two major evaluations of the Job Corps (Cain, 1968, and Mallar et al., 1980) and a third is now underway. Both studies have found very significant increases in employment and earnings, and reductions in welfare dependence, unemployment insurance usage, criminal activity, and out of wedlock births. Cost-benefit calculations indicate that for fiscal year 1977, the value of the benefits generated by Job Corps exceeded the costs of the program by almost 40% (Mallar et al.).

These Job Corps studies, however, are based on comparison groups rather than control groups. The absence of a randomly-assigned comparison group may
be especially critical for youth since it is not possible to rely heavily on preprogram earnings levels — as is often done with adults — to generate comparison groups.

The findings now emerging from the Youth Employment and Demonstration Act of 1977 (YEDPA) are, however, largely consistent with the Job Corps findings. A variety of approaches were used under YEDPA, but the most effective seemed to be a combination of intensive remedial education and training, along with training in job search techniques (Hahn and Ferman, 1984). Unfortunately, despite the fact that YEDPA was a demonstration, it was not run as an experiment. Consequently, even the most consistent findings are subject to the caveats mentioned above.

The National Supported Work Demonstration, run on an experimental basis, was designed to provide a highly structured and supportive work environment for four target groups: women with histories of long-term dependence on AFDC, ex-addicts, ex-offenders, and young school drop-outs. Of these four groups, the program was the most effective in improving the post-program (18 months to two years) employment and earnings of AFDC recipients, and had a substantial impact on ex-addicts as well.

Only for the AFDC recipients and ex-addicts do the benefit-cost calculations show that the program paid for itself on purely economic grounds. For AFDC recipients substantial increases in both hours worked and hourly wage rates were found, with accompanying decreases in receipt of welfare benefits. For ex-addicts, however, there seemed to be only small earnings gains, with most of the societal benefits coming from reduced criminal activity. These findings are not very controversial since they are, indeed, based on an experimental design.

One of the unexpected benefits of the Supported Work Demonstration is
that it has given us a capacity to check the estimated program impacts that have been generated via non-experimental methods against those of an experiment. In two separate analyses, Lalonde (1984) and Fraker, Maynard, and Nelson have generated "comparison groups" by methods similar, to those used in other studies. These comparison groups were then used to generate estimates of the net earnings impact of the Supported Work Demonstration which can then be compared to the "true" impact generated by the experiment's control group.

The results of these analyses are disquieting because the estimated impacts based on the comparison groups generally differed dramatically from the "true" impacts. Further, the comparison group results became progressively less reliable as the post-program period lengthened (Fraker, Maynard, and Nelson, 1984). Often the estimates based on the comparison group were statistically significant while those based on the control group were not, and vice versa. In general the comparison groups generated for youth were the least reliable while those generated for adult women were the most reliable.

The good news from these two studies is the confirmation that the comparison groups generated for adult women were probably the "best" of the available comparison groups. And the general finding that adult women benefit the most from participation is also confirmed by the Supported Work Demonstration. The bad news is that both the magnitude and the statistical reliability of even this "best" set of non-experimental results may badly mis-state the true effects. Clearly, the importance of an experimental design cannot be overstated.

An alternative to using public sector employment and training (which is essentially a 100% subsidy to state and local governments) is to subsidize wage costs for private employers when they hire workers from a particular target group. While this idea seems to have caught on in Europe, it has been
used on a very limited basis in the United States. Consequently, our ability to analyze the effectiveness of such programs is extremely limited.

The most recent of several wage subsidy programs is the Targeted Jobs Tax Credit. Under this program, companies can receive a tax credit on wages paid for certain groups of disadvantaged workers, including welfare recipients and youth from poor households. Employer participation in this program, as in its predecessor, the New Jobs Tax Credit, has been very limited. Many eligible individuals hold jobs with employers who do not claim the subsidy.

Why do firms and/or prospective employers fail to claim these tax credits? The results from a recent wage subsidy experiment suggest one possible explanation. In 1981 a Dayton, Ohio manpower agency randomly divided able-bodied welfare recipients into three groups. The first group was given a tax credit voucher informing prospective employers that the applicant was eligible for the Targeted Jobs Tax Credit, and outlining the precise nature of the credit. The second group was given a comparable wage subsidy voucher, this one to be reimbursed through a cash payment rather than a tax credit. The purpose of this was to test whether the low response rate was created by a fear of a possible tax audit associated with the credit. Both of these groups were encouraged to show the voucher when they applied for a job. The third group was not given any voucher.

In a recent analysis of the Dayton experiment, Burtless (forthcoming) found that twenty-one percent of the unvouchedered group found jobs within an eight week job search period, while only thirteen percent of the vouchedered groups found jobs. No significant differences between the two types of vouchers were found. Of the vouchedered job finders, only one quarter of their employers bothered to apply for the subsidy. This suggests that tax credit schemes may identify and stigmatize job applicants as being from a disadvantaged group.
This phenomenon of low response rates by employers (and perhaps prospective employees, as well) represents one of the most fundamental limitations of using the private sector to promote employment and training among very disadvantaged individuals. Although private sector programs may be much less expensive than public sector programs, they appear able to at best play a very limited role.\textsuperscript{10} This brings us to a discussion of the cost-effectiveness of employment and training policies.

The Relative Cost-Effectiveness of Alternative Programs

Given what appears to be the relatively modest effects of the programs, it is fair to ask whether or not employment and training programs for the disadvantaged represent a worthwhile expenditure of scarce resources. Do they generate more benefits either to the participant or society as a whole, than they cost?

Unfortunately, our inability to reliably estimate the benefits of employment and training programs makes any attempt to generate benefit-cost ratios extremely hazardous. Probably the most exhaustive benefit-cost ratios have been developed for the Job Corps and the Supported Work Demonstration.\textsuperscript{11} Although Job Corps is a very expensive program, it appears to almost pay for itself from the taxpayer's perspective — the program produces benefits of 96 cents for every dollar invested (Mallar, et.al, 1980). From a societal perspective (which includes both participants and taxpayers) the benefits are $1.45 for every dollar invested. The benefits to taxpayers come in the form of in-program output, increased tax payments on postprogram income, reduced transfer payments, reduced criminal activity, and reduced use of other federally provided services. If these estimates are accurate, very large benefits accrue to Job Corps participants at the cost of
a very small income redistribution from taxpayers.

The results from the Supported Work Demonstration, however, indicated that societal benefit-cost ratios for youth are well below one. However, it is impossible to know if the estimated difference is due to the difference in treatments or methodologies. Since the Supported Work results are based on an experimental design, it is difficult to dismiss them lightly. The results from Supported Work are, however, more positive for the AFDC target group, whose estimated benefit-cost ratios are well in excess of two: while those for ex-addicts are 1.87.

Benefit-cost calculations are always subject to criticism and controversy because of the heroic assumptions made to derive them. These calculations from the Job Corps program and the Supported Work Demonstration have, however been redone under a variety of assumptions and the basic conclusions stand. These results indicate that intensive training and employability development, like that found in these two programs, could be wise investments from society's point of view. Supported Work's benefit-cost ratios for ex-addicts and young school dropouts indicate that there is, however, no guarantee that such an investment will pay off for all disadvantaged groups and/or all types of programs.

There is much less evidence available on the societal benefit-cost ratios for more traditional (and less comprehensive) forms of employment and training such as those that existed under MDTA and CETA. However, a comparison of only the long-run increase in earnings resulting from various programs relative to their costs, suggests that training programs are a more cost-effective method of increasing disadvantaged participant's postprogram earnings than are employment programs. This result is not terribly surprising since the objective of many employment programs has been merely to provide temporary employment. However, employment programs (with perhaps the
exception of work experience programs) have probably been at least as effective as training programs in increasing post-program earnings. Employment programs, however, are much more expensive since participants are generally paid at least the minimum wage while training programs often at most only a small stipend.

From a benefit-cost point of view, in-program transfers to participants do not affect the societal calculations since the cost of the transfer to taxpayers is a benefit to participants. However, from the taxpayer's point of view, training programs are clearly preferable if they generate comparable earnings increases as do employment programs but at a lower cost. Also programs will be most cost-effective if they are tightly targeted on those groups that experience the largest earnings gains from participation.

**Fiscal Substitution and the Net Effects of Employment and Training Programs**

The extent to which employment and training programs can ultimately affect aggregate employment and earnings levels will be determined by a number of factors including: the number of individuals who participate, the impact that the programs have on participants, as well as their impacts on individuals who don't participate.

As Table 1 indicates, funding for programs grew steadily through the 1970s and then fell dramatically after 1979. Even at its peak, "average enrollments in CETA training components represented one-twentieth of the unemployed, less than a tenth of the low-income persons in the work force full-year and predominantly full-time but with earnings below the poverty level, and only half a percent of the labor force" (Taggart, 1981, p. viii). Clearly with the drastic decline in funding of structural programs since that
time, even these numbers have declined dramatically.

As the discussion up until this point should have made clear, employment and training programs have not been a panacea for disadvantaged workers. The long-run earnings gains that they generate seem to be (at best), fairly modest. These moderate earnings gains in combination with the low levels of funding, indicate that programs that grew out of the War on Poverty have certainly not eliminated, or probably even substantially reduced unemployment and poverty among the working age population.

But a complete evaluation of program effects must also consider the potential displacement effects that they generate. If the estimated employment and earnings gains that accrue to participants are at the expense of those individuals who did not participate in the programs, then the net macroeconomic effect of such programs would be zero.

Unfortunately, no formal analysis of this issue has been undertaken. Such an analysis would require a shift from the partial equilibrium frameworks used to evaluate the microeconomic impacts of the programs, to a general equilibrium framework that identifies economy-wide effects. Given the enormous difficulties that have been encountered with deriving defensible estimates of the microeconomic impacts, we seem a very long way from attempting this next step.

Some indirect evidence can, however, be derived. A number of studies estimated the "fiscal substitution" rate within the public sector. In this context, the term fiscal substitution refers to the extent to which state and local governments use funds from the federal government to finance expenditures on employment which would have been self-financed in the absence of the federal funding. The available estimates of fiscal substitution are based almost entirely on the public service employment (PSE) component of CETA.
Estimation of the fiscal substitution rate under PSE has been complicated by the fact that the program periodically underwent major changes. The first changes, in 1976, limited eligibility for PSE to the economically disadvantaged. The reauthorization of CETA in 1978 resulted in even further targeting of PSE on the economically disadvantaged. To the extent that substitution is less under more tightly targeted programs, we would expect the fiscal substitution rate under PSE to fall over time. On the other hand, as local governments adjust their employment levels to account for the extra employees provided by PSE, we would expect fiscal substitution to rise over time.

We will refer to the program prior to 1977 as "untargeted PSE," and the program after 1978 as "targeted PSE." The period from 1977-1978 represents a "transition" period from the untargeted to the targeted PSE. Several studies have attempted to measure fiscal substitution during these periods. Because of differences in methodologies, some estimate short-run effects while others present annualized (one year) or long-run effects (more than one year).

None of the studies has gone uncriticized and the confidence intervals around the estimates are either very wide or unreported. Our purpose here, however, is to look for a pattern in the results rather than exact orders of magnitude. The range of estimates from all of the studies is presented in Table 2. A pattern does emerge. In the shortrun, fiscal substitution is probably low but rises over time. Under the untargeted version of PSE, the long-run fiscal substitution rate was probably not significantly different from one hundred percent. Under the tightly targeted version of PSE, the long-run fiscal substitution rate is estimated to be between 0 and 29 percent.
Table 2
Estimated Fiscal Substitution Rate
by Duration of Funding

<table>
<thead>
<tr>
<th></th>
<th>Less than a Year</th>
<th>One Year</th>
<th>More than a Year</th>
</tr>
</thead>
</table>
| Untargeted PSE<sup>a</sup>  
(1974-1976)              | 0                | 50-60    | 100              |
| Transition PSE<sup>b</sup>  
(1977-1978)              | 18-21            | 36-70    | 23-70            |
| Targeted PSE<sup>c</sup>  
(1979)                    | 14               | 29       | 0-29             |

Note: These numbers represent percentages.

<sup>a</sup> These estimates are based on studies by Johnson and Tomola (1977) and Bassi and Fechter (1979).

<sup>b</sup> The short-run estimates are from studies by Cook (1980) and Nathan (1979), the intermediate estimates are from Bassi and Fechter (1979) and Adams-Cook-Maurice (1983), and the long-run estimates are from Adams (1981) and Bassi (1981).

<sup>c</sup> The short-run estimates are from Cook (1980), the intermediate estimates are from Adams-Cook-Maurice (1983), and the long-run estimates are from Adams (1981) and Bassi (1981).
There are, no doubt, a variety of factors that determine the rate of fiscal substitution. Some likely candidates are the state of the economy, the political environment, the level of PSE funding, as well as the restrictions which accompany the use of the funds. So to attribute the decline in the fiscal substitution rate to the restructuring of the program is to implicitly assume that fiscal substitution is unaffected by these other factors. Another shortcoming of these studies is that they only measure substitution within the public sector and do not attempt to measure any offsetting employment effects created in the private sector. Finally, these estimates tell us only about the potential displacement effects of employment programs but are silent about displacement resulting from training.

Despite the shortcomings of the fiscal substitution literature, one useful result emerges — employment and earnings effects will be overstated to the extent that the gains that accrue to participants represent losses for nonparticipants. The evidence presented in Table 2 suggests that this displacement (at least within the public sector) is likely to be less severe for structural programs than for countercyclical programs. An ironic implication is that structural programs may well generate larger employment effects than do countercyclical programs.

However, programs within the public sector that have low displacement/substitution effects may not be politically viable (Nathan, Cook, Rawlins, and associates, 1981). To the extent that local governments are not able to engage in fiscal substitution (since they typically hire very few low-skilled workers), they are not particularly interested in participating in the program. However, programs that have high substitution rates (and are, therefore, popular with local governments) are unpopular with unions. The elimination of PSE under the Reagan administration was not a difficult task since the program, by then, had very little political support.
If, indeed, this summary of events behind the collapse of large scale employment programs is correct, it seems unlikely that they can be counted on as a permanent method for increasing the employment and earnings capacities of the disadvantaged. Training programs are likely to be a more viable political alternative. And as was discussed in the previous section, they are also likely to be a more cost-effective vehicle for increasing earnings than are employment programs.

III. Summary and Conclusions

Judging from the evidence of the past, it is women and the economically disadvantaged who have benefited the most from employment and training programs. In most cases the resulting employment and earnings gains even to these groups have been modest, in part because it is not easy to solve the employment difficulties of the hard-to-employ, and in part because the resources devoted to any one individual are fairly modest. There is some indication that programs providing intensive (and expensive) investment in each participant, such as Job Corps and the Supported Work Demonstration have, at least for some groups of the disadvantaged, more than paid for themselves from a society wide point of view.

These findings suggest that employment and training programs have been neither an overwhelming success nor a complete failure in terms of their ability to increase the long-term employment and earnings of disadvantaged workers. Our ability to improve the lot of any given participant, and the collective economic well-being of the disadvantaged, has been modest — as has
been the level of resources that we have devoted to this endeavor.

These lessons have implications for the likely success of the Jobs Training Partnership Act which is still too young to be evaluated. JTPA has been very targeted on the most disadvantaged which, according to the evidence from the past, is the most effective targeting device for allocating scarce employment and training dollars. And because the program does not allow for any stipends to be paid to recipients, and is simultaneously to serve only the most disadvantaged, it is likely that only those on welfare will be able to "afford" to participate in the program. This essentially limits the program to welfare mothers, which is precisely the group that is most likely to benefit from the program.

Unfortunately, the level of support that is available to any given participant remains limited as no stipends are available and maximum length of program participation is limited. Given the severe employment barriers that many participants face, we may be making a mistake which we have made before -- investing too little and hoping for too much.

Unfortunately, despite nearly twenty years of continuous federal involvement, we still have to do a good deal of guess work about what will work and for whom. We have had substantial, and on-going difficulties, in identifying what works - much less for whom, and why it works. This has been in large part, because of an unwillingness on the part of Congress and policy makers to allow for adequate experimentation in the delivery of employment and training services. As long as analysts are forced to use comparison groups, instead of true control groups, there will always be debate and controversy over their effectiveness.

Two decades of non-experimental program evaluation in the employment and training field have finally taught a lesson about which there can be little disagreement: Convincing program evaluation is going to require continued use
of randomized clinical trials. We wish to emphasize that this is not simply a statement that randomization is a preferable methodological approach regardless of the field of study. Instead, we believe the evidence in the study of employment and training programs overwhelmingly indicates that randomization is essential in program evaluation in the employment and training field. The difficulty seems to be that the earnings determination processes of today's workers and the program selection methods of today's programs interact to make it nearly impossible to produce reliable estimates of how workers' earnings would have behaved in the absence of a program. The evidence to support this finding comes from both experimental and non-experimental studies. The non-experimental studies indicate that minor changes in methods, for which there is no empirical justification, produce large swings in estimated program effects. The study of experimental findings indicates that perfectly plausible non-experimental methods may lead to dramatic errors in inferences about program effects.

This basic finding raises a fundamental question: What is the proper reaction of policymakers? In our view the appropriate reaction is for policymakers to begin the development of a credible research and development effort using randomized clinical trials in a wide variety of study areas. It is even possible that enough may be learned from this approach in a decade or two of experience that non-experimental analyses will come back into a larger role. We wish to emphasize, however, that it will take at least a decade before the full fruits of the effort will be in evidence. The successful use of randomized clinical trials in the Supported Work Experiment indicates that, like the great progress at the Food and Drug Administration and the Federal Trade Commission in the use of statistical methods, randomized clinical trials is an entirely practical and ethical way to deal with model uncertainty in the employment and training area.
We believe that experimentation will naturally proceed in two types of areas, and that whether one or the other is favored will no doubt depend in part on the political environment. One kind of experiment continually tests and evaluates on-going programs, while the other investigates prototypes for new programs. Building up a treatment and control comparison in the longitudinal analysis of the Job Training Partnership Act is an example of the former. There is no doubt that a similar approach could be applied to the Job Corps, the Neighborhood Youth Corps, or any of the other large scale employment and training that now exists.

More challenging in some ways is the use of experimentation in the design of new programs or the modification of continuing programs, such as the unemployment insurance system, the disability income system, or other welfare programs. Here small scale modifications can be treated by randomized clinical trials in selected sites. Likewise, the development of new programs can be tested in selected sites before a program on a national scale is implemented.

It will not be until we have accumulated evidence from this type of experimentation that even the most consistent finding from the evaluation literature — that women benefit the most from program participation — can be accepted without doubts. There is always the possibility that women participants appear to be more successful in the post-program period because those who have chosen to enter or re-enter the labor force self-select into employment and training programs, and those women who have chosen to remain out of the labor force are over-represented in the comparison groups. Of course, to the extent that these programs facilitate entry or re-entry, the estimated program impacts are "real." But to the extent that the programs simply are a vehicle for entry, for those who would have entered even without the program, the estimated program effects will be overstated. No selection
bias adjustment mechanism can promise to eliminate the effects of this type of nonrandom assignment.
A more detailed set of results are available from the authors on request.

Although it is impossible to correct for this problem, it is possible to estimate the magnitude of the bias that it introduces in the estimates.

Ashenfelter's main estimation technique was an autoregressive earnings model, where lagged earnings was used to control for differences in unobservable characteristics between participants and comparisons.

Sample sizes vary from year to year — with the smallest consisting of 6,700 participants during 1975. In later years, the samples were in excess of 15,000.

This discussion is based on the work of Bassi (1982a, 1984), Bassi et al. (1984), Bloom and McLaughlin (1982), Dickinson Johnson and West (1984), and Westat (1979).

It should be noted, however, that Bassi et al. (1984), found very little evidence of significant program effects for male youths (age 16-22).

The Bloom-McLaughlin (1982) evaluation was the only study not reaching this conclusion.

Significant program effects were not found for young women with children.
Lalonde (1984) generated comparison groups from both the PSID and CPS while Fraker, Maynard, and Nelson (1984) relied exclusively on the CPS.

Private sector programs that involve wage payments are less expensive than comparable public sector programs since public sector programs use a 100% wage subsidy while private sector programs use a subsidy of less than 100%.


This statement would not necessarily be true if income redistribution was an objective of the program.

In this section we confine our attention to the employment effects of these policies and do not consider inflationary effects.

Note that the concept of fiscal substitution is slightly different from that of displacement. Fiscal substitution measures a substitution of money, while displacement measures substitution rates among workers. We would expect, however, that the two measures would be highly correlated.

This assumes that the effects of the reauthorization were not immediate but occurred in the year following the legislative change. A Congressional Budget Office (1981) report shows that, in practice, PSE did not become tightly targeted until after the legislative changes in 1978. During 1977-1978, the program was only slightly more targeted than it was in 1976.
References


