SOCIAL POLICY RESEARCH ASSOCIATES

What's Known About the Effects of Publicly-Funded Employment and Training Programs

Final Report
April 21, 2006

Prepared by:

Ronald D'Amico

Prepared for:

U.S. Department of Labor
Employment and Training Administration
200 Constitution Ave., N.W.
Washington, D.C. 20210

DOL Contract AF-12526-02-30
Project No. 1337

This project has been funded, either wholly or in part, with Federal funds from the Department of Labor, Employment and Training Administration under Contract Number AF-12526-02-30. The contents of this publication do not necessarily reflect the views or policies of the Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

1330 Broadway, Suite 1426
Oakland, CA 94612
Tel: (510) 763-1499
Fax: (510) 763-1599
www.spra.com

ACKNOWLEDGEMENTS

This report was prepared under contract to the U.S. Department of Labor, Division of System Accomplishments and Accountability. The author is grateful for the helpful comments provided by Daniel Ryan, of DOL; Professor David H. Greenberg, of the Department of Economics at the University of Maryland; Dr. Christopher T. King, Director of the Ray Marshall Center at the University of Texas at Austin; and Richard West, President of SPR. Professor Greenberg and his colleagues, Charles Michalopoulos and Philip Robins, also generously shared some of the raw data on which they based their recently completed meta-analysis of employment and training programs, which was very helpful in the preparation of this report.

CONTENTS

WHAT'S KNOWN ABOUT THE EFFECTS OF PUBLICLY-FUNDED EMPLOYMENT AND TRAINING PROGRAMS

Federal appropriations for the U.S. Department of Labor's (DOL's) workforce investment programs amounted to nearly $9 billion for 2005, encompassing a diverse array of programs authorized under the Workforce Investment Act and related legislation (U.S. Department of Labor, 2005). Although smaller in real dollar terms than it was at its peak in the early 1980s (King, 2004) and modest as a percentage of GDP in comparison to many other Western nations (OECD, 2001), this amount nonetheless reflects the nation's concerted and long-standing commitment to active labor market policies designed to help those who are struggling attain a firmer foothold in the labor market.

In an effort to ensure that these resources are well spent, DOL and other funders have also supported rigorous experimental and quasi-experimental evaluations over the past several decades of what works and for whom. By now, a substantial body of literature has accumulated on impacts of diverse types of publicly-funded employment and training programs for various participant populations. Unfortunately, these sometimes yield divergent assessments. Indeed, even evaluations of the same programs using largely the same data have sometimes produced widely discrepant impact estimates (e.g., Barnow, 1987). Moreover, the record is still woefully incomplete in many respects. For example, we know little about how the quality of program implementation across sites affects program impacts.[1] Thus, any conclusions about what works and for whom are subject to substantial remaining uncertainty.

Despite this, a consensus of sorts has emerged, which forms the basis for this paper. Moreover, this consensus seems to be supported whether one looks at results from quasi-experimental or experimental evaluations; indeed, although the former may produce estimates that are more variable, the same general conclusions seem to hold (LaLonde, 1995; Heckman et al., 1999; Greenberg et al. 2005a).

[1] Results from the Minority Female Single Parent Demonstration (Burghardt et al., 1992) suggest that how a program is operated can have important implications for how successful it is. See also Bloom et al. (2003).

We review this evidence in this paper, focusing on the effects of *publicly-funded* employment and training programs *in the U.S.*, primarily targeted to economically disadvantaged out-of-school youth and adults.[2] We first discuss some of the methodological issues involved in estimating program impacts. We next report the array of program impacts that have been estimated for various target populations and service strategies, the rate of decay (or, conversely, durability) of program impacts in the postprogram period, and whether short-term performance indicators provide any clue as to whether a program will be shown to be effective in the longer-term. Our focus is on summarizing program impacts with respect to annual earnings. In general, we concentrate on reviewing the literature pertaining to voluntary employment and training programs, although some discussion of the impact of mandatory programs is provided as well.[3]

For reference, Appendix A provides a brief description of the key features of many of the programs we review, and Appendix B provides an annotated bibliography.

Methodological Issues

Although documenting the *outcomes* associated with participation in employment and training programs is relatively straightforward, estimating program *impacts* is altogether more challenging for a number of reasons. Most importantly, the latter implies some knowledge of the counterfactual—specifically, how the outcomes achieved by program participants compare to what they would have been in the absence of the intervention. Additionally, in practice evaluation research has been limited in the inferences it can make because of the highly variegated nature of the interventions that are typically being studied and the limited range of outcomes and time periods being examined. Each of these issues is discussed in turn.

---

[2] In dollar terms, employer-funded training appears to be orders of magnitude more prevalent than publicly-funded training (see, for example, Lerman et al., 2004), and its effects in boosting workers' earnings have received some treatment in the empirical literature (e.g., Lillard and Tan, 1992). And, of course, the returns to formal schooling have also been much researched. Neither of these topics is treated in this paper, except in passing. The U.S. has a much longer tradition of conducting rigorous evaluations of publicly-funded employment and training programs than other nations, but some research is emerging on the international experience (see Heckman et al., 1999, for a review of this literature).

[3] It is useful to make a distinction between mandatory and voluntary programs because the two can have effects in distinctly different ways. With respect to voluntary programs, one assumes that any impact on outcomes (e.g., employment and earnings) comes about because the services are efficacious. Mandatory programs, by contrast, can generate effects for this reason, as well as because the compulsory nature of the intervention means that some of those covered by the mandate to participate could change their behavior in ways that influence outcomes whether or not they actually access program services. For example, persons could accelerate their job search and accept employment simply to avoid having to participate in a mandatory job-search workshop. In this way, program impacts on those not actually receiving services can come about.

Dealing with the Counterfactual

When generating an estimate of program impacts one wants to know how the outcomes attained by participants differ from what they would have been if the individuals had not participated. The latter obviously cannot be directly observed and is thus a counterfactual.

Unfortunately, taking into account the counterfactual in developing impact estimates is particularly problematic for two, inter-related reasons. First, program participants' immediate pre-participation work history is rarely a reliable guide as to what their post-participation employment experiences would be in the absence of training. One important reason for this is that participants experience a well-known *earnings dip* before program participation begins—presumably part of the reason they seek out services to begin with (e.g., Ashenfelter, 1978). In pre-post comparisons, whether one judges a program to be effective or not depends mightily on whether one thinks the pre-program earnings dip in the absence of services is transitory, and, if so, how quickly one expects earnings to rebound and to what level. For this reason, simple pre-post comparisons generally are not judged to be a satisfactory means for assessing program effectiveness, giving rise to the search for a comparison group of persons not served whose outcomes may serve as a proxy for what the outcomes of program participants would have been in the absence of program services.

However, identifying an appropriate comparison group to be used in this way is itself problematic because of the second key reason that makes treatment of the counterfactual so difficult. Specifically, the employment and training programs that are the subject of this review are voluntary in nature and are (generally) not entitlement programs. Thus, persons who seek program services do so of their own volition; further, of those seeking services, some might be denied enrollment if program administrators judge them unsuitable for the services that are available. The joint participant-administrator decisions—the prospective participant's decision to seek services and the program administrator's decision to enroll a subset of those who apply—thus give rise to potentially grave *selectivity issues*, such that those who are enrolled cannot readily be compared to non-enrollees for purposes of deducing program impacts. For example, among those who experience a similar pre-program earnings dip, program applicants might be more highly motivated to succeed than those who do not seek services, and, hence, could be expected to do better whether or not they obtained program services; alternatively, they might judge their employment prospects so poor on their own that they seek out services, whereas others who are similarly situated might engage in concerted job search on their own, confident of their employment prospects even without assistance. Similarly, program administrators might give preference to enrolling applicants whom they judge most likely to succeed (for example, to improve their program's prospects for meeting its performance standards), or they might give

preference to the very hardest to serve, on the supposition that these are the ones who need program services the most.[4] In short, selectivity issues can be assumed to come into play, but the direction of any potential bias cannot be known with any certainty in advance.

Quasi-experimental methods were typically used to address selectivity bias in evaluations of employment and training programs that were conducted several decades ago, when concerted efforts to identify program impacts commenced. Typically, a comparison group drawn from a national data base (e.g., the Current Population Survey) was identified by using statistical matching methods, such that comparison group members would be like their treatment group counterparts on an array of measurable attributes, including their pre-program work history and demographics (see Dickinson et al., 1986, and Bassi, 1984, for examples). Typically, econometric modeling was also used, to purge the estimated treatment effect from the influence of as many potentially confounding factors as possible.

Unfortunately, reviews of this literature pointed out that studies of the very same program and using the same data could sometimes produce widely divergent estimates of program impact depending on how the comparison group was selected and precisely how the estimation model was specified (see, for example, Barnow, 1987), casting widespread doubt on the appropriateness of quasi-experimental methods in this context. Indeed, looking at this evidence, DOL's Job Training Longitudinal Survey Research Advisory Panel (Bloom et al., 1985) advocated that DOL rely on experimental methods in evaluating its programs, and partly for this reason for at least the next decade experimental methods generally held sway.

The strong allure of an experimental design is that, by randomly selecting a treatment group and a control group from among eligible applicants, one is guaranteed that treatment and control group members are identical on both observables and unobservables, except insofar as they differ purely by chance. In other words, the problem of a systematic selectivity bias can be entirely and convincingly addressed.

At the same time, it was soon discovered that, in practice, experimental methods have limitations of their own. For example:

- Experiments are typically extremely costly to run.
- They are very difficult to implement in a real-world context. To begin with, program administrators are often reluctant to cooperate if it means they must deny services to eligible applicants who are assigned to the control group. Similarly,

[4] See Heckman et al. (2002) for some evidence on whether program administrators are motivated more by factors that induce them to "cream" or to serve the hardest to serve.

effectively running an experiment in multiple sites is highly burdensome. For these reasons, experiments are often carried out in just a few sites or ones that may not be randomly chosen, causing the findings to have questionable external generalizability.

- Experiments often yield summary judgements of whether a program works or not, but cannot readily shed light on what combinations of services work best and in what context. This is because statistical equivalency between the treatment and control groups is generally lost when one begins to parse out the treatment group into substrata defined on the basis of program services that participants actually received. For this reason, experiments are often viewed as "black box" evaluations.[5]
- In some circumstances an experiment simply cannot be carried out, such as when as a matter of law all eligibles are entitled to receive program services if they apply.[6]

For these reasons, quasi-experimental methods still retain their place, a point that has been compellingly argued of late by Heckman and Smith (1995). Further, recent reviews of the literature have concluded that quasi-experimental methods yield impact estimates that are on average no different from those generated through experimental designs and lead to similar policy conclusions (e.g., Greenberg et al., 2005a; LaLonde, 1995; Heckman et al., 1999).[7] Moreover, although it can be futile to try to gauge whether or to what degree quasi-experimental estimates might be biased even once we know how the study was designed (e.g., Glazerman et al., 2002), in recent years we have come to understand much better how to carry out quasi-experiments that yield credible impact estimates.[8] As summarized by Heckman et al. (1999) and Bloom et al. (2002), these key conditions are that the quasi-experimental study should:

- Compare treatment and comparison group members who are drawn from the same local labor market.

---

[5] The JTPA Experiment represents an important though only partially successful effort to overcome this limitation, by conducting the randomization after eligible applicants were assigned to services (Bloom et al., 1994).

[6] For example, the Trade Adjustment Assistance program is an entitlement, in that all those who experience a qualifying, trade-related separation are entitled to the benefit of program services as a matter of law. For this reason, SPR was contracted by DOL to carry out a quasi-experimental, rather than experimental, evaluation of TAA (Social Policy Research Associates, 2005).

[7] Greenberg et al. (2005a) base their judgement on a meta-analysis of the effects of employment and training programs estimated over the past three decades and conclude that whether an impact estimate was generated through experimental or quasi-experimental methods seems to make no significant difference.

[8] For example, some researchers (e.g., Heckman et al., 1997) have attempted to replicate experimental findings for a single program using quasi-experimental methods, to learn under what circumstances the latter can closely replicate the former.

- Measure variables from the same data sources for both groups, and
- Have a rich array of variables to use as control variables or for modeling the participation decision, including information on pre-program work history.

Thus, whether quasi-experimental methods would be considered as a viable option in a given context depends on whether experimental methods are deemed suitable (e.g., whether carrying out an experiment is feasible, given the research budget, program administrators' receptiveness to randomization, and the research questions under investigation) and whether the quasi-experimental glass is considered half-full or half-empty—the possibility that estimates of program impact are contaminated by selectivity bias can never be entirely ruled out under a quasi-experimental design, but the size of the bias appears to be close to zero on average. Moreover, given rich data and appropriate methods the risks of bias can be minimized while allowing for the investigation of a fuller array of research questions that are theoretically interesting and highly policy relevant.

Other Methodological Issues

Apart from how the problem of selectivity bias is addressed, another methodological challenge associated with estimating program impacts and drawing implications for policy stems from the *varied nature of the intervention* itself. In contrast to estimating treatment effects in medical sciences, for example, where the intervention will typically represent a carefully specified dose of a carefully formulated substance, interventions in employment and training programs are often highly variable. For example, depending on the customers' perceived needs and the preferences of the program operator, a participant served under the Workforce Investment Act might be given job search assistance, prevocational training, classroom training, workplace training, or work experience, provided either alone or in combination, and accompanied by varying degrees of case management, supportive services, assessment and counseling. Each service in turn could be delivered differently and for different durations and intensities (for example, publicly-funded classroom training can provide participants with instruction in a variety of technical fields and last a few weeks, a few months, or several years). As a consequence, different individuals served by the same program could thus effectively receive very different interventions. Moreover, different program operators running the same program in different sites (e.g., different local workforce areas operating WIA programs) will likely have different service emphases and designs, potentially causing pronounced site-to-site variation in what services are typically provided and how. Given all this variability, in evaluating whether "it" is effective or not, it often is not entirely clear what the "it" is that is being studied.

Drawing policy implications about the effectiveness of an intervention so variously defined is further complicated because some members of the control group will access very similar services

elsewhere, while some treatment group members will drop out before receiving anything. In this sense, the counterfactual has to be clearly understood as the incremental effect of *gaining access to a given program's* services, and should not be interpreted as the effect of receiving services versus receiving no services at all. These combined issues of so-called dropout and substitution bias (Heckman et al., 2000) clearly cloud the conclusions we can safely draw from evaluation efforts, as we will discuss more fully later in this paper.

Still another issue in judging whether an intervention is effective relates to *what outcomes should be measured and when*. In principle, participants of job training programs could reap any number of benefits from participation, both direct and indirect, including improvements to their work skills or the attainment of a credential, or changes to employment probabilities, hours worked per week, hourly wages, earnings, unemployment or job turnover rates, job satisfaction, and public assistance recipiency, among other things. Further, changes to these outcomes brought about by the program could vary greatly depending on whether they are measured while services are being provided, shortly after they are concluded, or much later. For example, earnings will usually be depressed during the period of program participation, because participants will often drop out of the labor force or scale back their work hours while they are pursuing training. The expectation is that their earnings will then rebound when their training ends, such that they catch up to, and eventually surpass, their comparison group counterparts. Thereafter, program participants could maintain their earnings advantage or not, depending on whether the intervention imparts a long-term advantage in the labor market or its effect eventually wears off. In other words, our assessment of whether the intervention was effective could vary, depending on what outcome we choose to look at, the program's inherent long-term efficacy, and whether we measure its effects one year after the treatment begins, or two years, five years, or ten years after.

In actuality, the metric of choice for measuring program impacts is typically earnings, because it has been viewed as an appropriate summary measure for evaluating the effectiveness of workforce program and can be measured fairly reliably through administrative records (such as Unemployment Insurance wage records). Further, given the costs of data collection and analysis, outcomes are rarely measured for more than a few years after participation ends. Thus, despite the potential for using longitudinal microdata to examine a wider range of outcomes over a longer period of time, in general analysts have based their decisions about whether one program is more effective than another by observing impacts on a narrow range of outcomes (such as earnings) in the relatively short-term.

A Summary of Program Impacts

Notwithstanding the problems mentioned above, analysts have estimated program impacts for various participant populations and for programs of different types, and have assessed how long effects seem to persist. Each of these topics is discussed in turn. Because, as we have discussed, quasi-experimental methods in general appear to yield reliable conclusions, we draw on findings from both quasi-experimental and experimental evaluations, to expand the range of information on which we base conclusions. Our focus is on impacts on annual earnings, because, as discussed, this is the metric that is uniformly available.

Impacts for Target Populations

Researchers have tended to report evaluation findings separately for at least four major target groups—adult women, adult men, and youth who are economically disadvantaged, and dislocated workers (though research on the latter group is quite sparse). Evaluation findings are broken out this way partly because some programs are targeted to one of these target groups but not the others (e.g., Job Corps serves only youth and young adults, and the Minority Female Single Parent Demonstration served only adult women), and because early experience has shown that each group tends to fare differently as a consequence of program participation.

Exhibit 1 shows the distribution of program effects for the first three of these four groups (there are too few separate program impacts for dislocated workers for this group to be meaningfully added to the chart).[9] The exhibit makes two key points abundantly clear. First, there is substantial variability in estimated earnings impacts within each of the three groups. Thus, for each group, some earnings impacts are estimated to be negligible or even negative while others are strongly positive, amounting to $2,000 or more per year. Consistent with a point made earlier in this paper, this variability likely reflects not only some measurement error but the fact

---

[9] The table was assembled by compiling the reported impact on annual earnings from the many separate studies published over the past several decades, but restricted to those where the services were provided in 1975 or later. We compiled data directly from the original sources for many studies, but much additional data was gleaned from a database generously provided by Greenberg, Michalopoulos, and Robins, on which they based their recently completed meta-analysis of employment and training programs (Greenberg et al., 2003a). For consistency, all effects were inflated to 2003 dollars and represent the estimated earnings impact in the second full year after the year in which services were provided. Effect sizes were inflated to 2003 dollars using the Consumer Price Index. The second year after the year in which the training commenced was chosen as the focus of the exhibit because this period gives participants the chance to complete their training and reap labor market returns from it; it also represents among the latest years for which impact estimates are reported in most studies and is the approximate time during which earnings impacts of employment and training programs appear to reach their peak (as we will show later in this paper). The compilation was somewhat arbitrarily restricted to studies of services that were provided in 1975 or later, because impact estimates from earlier studies (primarily MDTA) are deemed to be less reliable.

that the impacts estimated for a single group represent the effects of interventions that are themselves very different and which are examined in different local contexts. In short, there is no consistent answer as to how much publicly-funded employment and training programs help boost the earnings of program participants.

The second point brought out by the exhibit is that, notwithstanding the intra-group variability in effect sizes, some groups clearly make out much better than others do, at least on average. For example, adult women seem to fare the best, as almost none of the impacts estimated for them are negative and relatively few are positive but negligible in magnitude, but many are positive and sizable. In fact, about a third of the effect sizes show impacts on annual earnings for women of $2,000 or more.

At the other extreme, youth seem to fare least well, and in fact seem to be hurt as often as helped by the services that are provided to them, at least insofar as impacts on annual earnings are concerned. Thus, about half of the estimated earnings impacts are negative and most of the others, although positive, are quite small, at less than $500 per year. (Of course, because our focus is on employment programs, we are making no judgment about the efficacy of programs providing academic instruction to in-school youth, which, if they are successful in boosting youths' academic achievement, could lead to earnings gains many years later.)

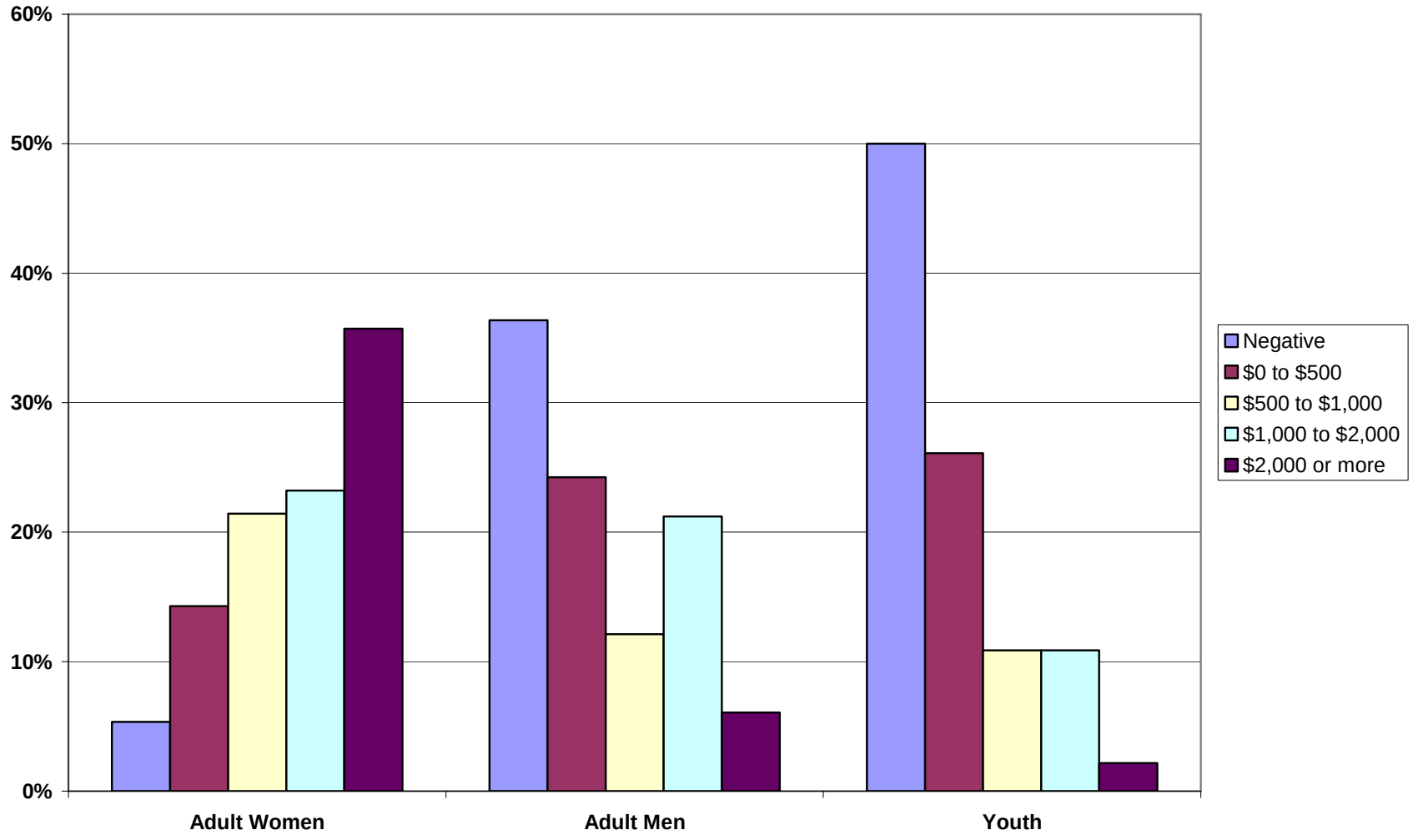
Finally, the effects for adult men appear to be in between these extremes. Adult men realize negative earnings impacts in many more studies than adult women do, but still typically experience positive earnings impacts, and often times ones that are sizable.

In this section we consider each of these groups in somewhat more detail. We also discuss findings that have been reported for dislocated workers, which, although too scant to be included in the exhibit's histogram, nonetheless provide clues as to how this important target population fares.

Economically Disadvantaged Adult Women

Program impacts seem clearest for economically disadvantaged adult women. Across interventions of a variety of types, findings are consistently more favorable for this target population than for any other group—impacts estimated for them are more likely to be statistically significant, are larger in magnitude, and seem to persist the longest. Significant and positive average impacts emerged in the days of MDTA (e.g., Ashenfelter, 1978), continued through CETA (e.g., Bassi, 1983 and 1984) and were found again in JTPA (Bloom et al., 1994 and 1997). Impressive positive earnings impacts surfaced as well in special demonstration

Exhibit 1: Distribution of Annual Earnings Impacts, by Group



programs providing long-term supported work experience, such as in the National Supported Work Demonstration (Couch, 1992) and the Homemaker-Home Health Aide Demonstration (Bell and Orr, 1994). Moreover, interventions seem at least as efficacious, and perhaps even more so, for the much studied group of adult female welfare recipients (see Bloom et al., 1997; Couch, 1992; Nudelman, 2000), a subset of special policy interest in light of recent welfare reforms. The one notable exception to this generally very positive picture is the disappointing findings from the Minority Female Single Parent Demonstration (MFSPD), which failed to find significant effects on employment and earnings from this effort to provide a sequence of basic skills remediation, followed by job-skills training, to single mothers (Burghardt et al., 1992), possibly because substantial numbers of those served failed to complete the basic skills training to which they were assigned (Hollister, 1990) and thus never moved on to job skills training (Burghardt and Gordon, 1990).

In looking at the weight of this evidence, then, it can be concluded that voluntary programs for adult women are generally effective in a wide variety of contexts (Friedlander et al., 1997; King, 2004; Grubb, 1996; LaLonde, 1995). Greenberg and his colleagues (Greenberg et al., 2003a) reach the same conclusion in their formal meta-analysis of 315 impact estimates. Moreover, although the evidence is scant, programs seem to be effective for economically disadvantaged women more or less regardless of their different baseline characteristics. For example, we have already noted that welfare women benefited about as well from JTPA as the non-welfare group, and Bloom et al. (1994) find no evidence that other potential barriers to employment are related to estimated program impacts in any systematic way.

Economically Disadvantaged Adult Men

As Exhibit 1 showed, findings are a bit less clear for adult men, as impacts estimated for them can vary widely. The studies of MDTA (e.g., Ashenfelter, 1978) showed significant positive earnings impacts for men, but with effects that were generally smaller than they were for women. CETA findings for this target group are quite diverse, ranging from large and negative (e.g., Dickinson et al., 1986 and 1987) to large and positive (Bassi, 1983), depending on the estimation method used and the CETA service strand being considered. The findings from the more recent JTPA experiment show consistently positive effects for men, but again the earnings boost they realize appears to be more modest than for women. There have not been as many special demonstration projects for men as there have been for women, or at least ones subject to critical empirical examination, but the most positive effects for men come from the one exception, the National Supported Work Demonstration, which served ex-offenders and ex-addicts and showed substantial earnings impacts for them.

In sorting through these findings, it appears necessary to conclude that earnings effects are consistently more modest for adult men than for adult women (e.g., Grubb, 1996; Friedlander et al., 1997; King, 2004), but that, on balance, men do appear to realize a net gain from program participation (Greenberg et al., 2003a).

Economically Disadvantaged Youth

As Exhibit 1 has suggested, employment and earnings impacts are consistently less favorable—in fact, are often unfavorable—for employment programs targeting young adults. This is true for both young men and young women and virtually regardless of the program being studied, a conclusion that is now widely shared (Lerman, 2000; LaLonde, 1995; Grubb, 1996; Friedlander et al., 1997; Greenberg et al., 2003a).

For example, work experience programs for teens that have been a staple of youth programming for many years have been evaluated in several incarnations, and none has been shown to be effective beyond the very short-term; these programs include the National Supported Work Demonstration (Couch, 1992), STEP (Grossman and Sipe, 1992), and the Summer Career Exploration Program (McClanahan et al., 2004). In considering a wider range of interventions, the JTPA evaluation found that young adults reaped virtually no significant gains from program participation from any activity in any year (e.g., Bloom et al., 1997; U.S. General Accounting Office, 1996; Westat, nd). Even very costly and intensive interventions providing access to a comprehensive array of services over a long period of time, such as JOBSTART (Cave et al., 1993), New Chance (Quint et al., 1997), and Project Redirection (Polit et al., 1985), do not appear to have yielded significant earnings gains. Optimists who believe, with some justification, that the JTPA experimental results for youth have been interpreted too harshly (e.g., Krueger, 2003) must find this overall picture sobering.[10]

There has been some scattered good news, however. Although JOBSTART failed to produce significant post-program earnings gains overall, one of the thirteen sites in which this program operated, CET in San Jose, proved to be an exception, showing significant earnings gains in the third year after random assignment (Cave et al., 1993). Interestingly, CET also was the only one of four sites that produced significant earnings gains for single mothers, as part of the Minority Female Single Parent Demonstration (Zambrowski and Gordon, 1993), fueling speculation that

---

[10] Our review is of programs whose primary aim is to improve post-program employment and earnings. Thus, we generally exclude from consideration the many youth programs that focused primarily on improving school retention or academic achievement, even though youth programs that improve academic standing can be presumed to (or have been shown to) improve employment outcomes. Programs of this sort include Ohio's LEAP (Bos and Fellerath, 1997) and the Quantum Opportunity Program (Shirm and Rodriquez-Planas, 2004).

CET's model of integrated basic and occupational skills instruction in a work-like setting could serve as a successful model for youth programming.

Based on these positive results, DOL funded the CET Replication study, designed to examine whether the CET model could be successfully replicated and could generate positive impacts on the employment and earnings of young adults elsewhere. Researchers found that the CET model was in fact difficult to replicate, but that, even where replicated well, there were no significant and positive effects on earnings for either young men or young women (Miller et al., 2003).

Results from the National Job Corps Study are somewhat more encouraging, though still mixed. Enthusiasm first surfaced as a result of the encouraging findings from the quasi-experimental evaluation of Job Corps that was conducted more than two decades ago (Mallar et al., 1982). Initial findings from the more recent experimental evaluation were equally positive (Schochet et al., 2001). As would be expected for those enrolled in an intensive training program, Job Corps members had substantially lower earnings in the program period, but their earnings rebounded sharply upon exit from the program, showing an earnings impact that persisted through the end of the initial four-year follow-up period. However, more recent analyses using administrative data for the fifth through ninth years after randomization severely temper this conclusion, as they show that the impact of Job Corps shrinks to near-zero overall during the fifth year after randomization and remains there through year nine—very disappointing news for a program as costly as Job Corps. An exception, though, is for older youth (those ages 20 to 24 at program application), who retain a sizable earnings advantage, even through the ninth year (Schochet et al., 2003 and 2005). Thus, Job Corps appears to be efficacious even in the longer-term for older enrollees, but not for those who are younger.[11]

Dislocated Workers

Little rigorous research has been conducted on the effectiveness of programs to help dislocated workers rebound from the sizable earnings losses that they usually encounter following their job separation (Jacobson et al., 1993). Some evidence, though, comes from the Worker Adjustment Demonstrations, funded by DOL in the early 1980s, including projects in Downriver (Michigan), Buffalo (New York), and Texas (three sites). These projects offered job search assistance, as well as retraining services for some. Leigh, who conducted a careful review of this literature, concluded that job search assistance seemed to have short-run impacts on earnings, but that

[11] Note that this is an age group older than that typically covered by most other youth programs being evaluated here. Thus, it is questionable whether the Job Corps findings for those in this age range can be counted as an exception to the generally poor results for youth programs.

retraining services failed to boost earnings beyond what occurred with job search assistance alone, yet was much more costly (Leigh, 1990 and 2000; see also Kodrzycki, 1997).

An obvious deficiency with these studies, though, is that the follow-up period was quite short—only 2.5 years from the date of layoff in the case of the Downriver Demonstration and no more than 1 year after intake in the remaining two projects—arguably too short to see much of an effect from a retraining investment.[12] The six years of follow-up afforded by the New Jersey Unemployment Insurance Demonstration Project provide a more appropriate time horizon in which to evaluate a similar intervention. This mandatory program assigned dislocated workers who were UI claimants to one of three treatment groups: those receiving only job search assistance, those receiving job search assistance plus the offer of retraining, and those receiving job search assistance along with a reemployment bonus. One year after the initial claim, those receiving only job search assistance recorded earnings gains that were as large as those who were also offered retraining (Corson et al., 1989). Additional data from a six-year follow-up showed that the earnings impact persisted for all three groups, but, again, those who were offered retraining seemed no better off than those who received only job search assistance (Corson and Haimson, 1995). However, in actuality only 15% of those offered training actually received it, so the failure to find an additional earnings impact for this group should not be surprising. Indeed, supplementary regression analysis leads the authors to conclude that in fact retraining did boost earnings somewhat among those who received it.

Additional evidence on the effect of retraining is also mixed. In their study of dislocated workers who received training funded under the Trade Adjustment Assistance Program, Corson et al. (1993) fail to find any significant effects of training. On the other hand, Jacobson and colleagues (Jacobson et al., 2002, 2003 and 2005) use data on displaced workers from Washington State to conclude that those who complete technical courses at community colleges realize a substantial earnings boost, but one not shared by those who complete courses in less technical fields.

In short, evidence thus far suggests that job search assistance can be effective and can have effects that persist for dislocated workers, but the added advantage bestowed by retraining is unclear. However, the limitations of the available studies preclude a firm judgement.

---

[12] A second problem that plagued at least the lone Texas site where training was available as an option was that the training offered at the local community college appears to have been badly matched to the workers' interests, causing extremely low training take-up rates (Leigh, 1990).

Impacts by Program Type

Given the evidence just discussed, it is important to examine more systematically whether some types of interventions appear to work better than others. Thus, we consider relatively light-touch services, like job search assistance, and those providing more substantial skill building.

Job Search Services and Placement Assistance

As we have noted, job search assistance seems to be effective for dislocated workers, but it has been shown to be useful in other program contexts and for other target population as well. For example, in addition to the work on dislocated workers just cited, a similarly impressive body of work reaches similar conclusions for the broader population of the unemployed (O'Leary, 2004). In particular, job referrals from the Employment Service have been shown to appreciably shorten the duration of unemployment (e.g., Jacobson and Petta, 2000), especially for females (Johnson et al., 1985).

Other assessments of the efficacy of job search services are drawn from studies of mandatory programs, targeted to either Unemployment Insurance claimants or welfare recipients. Meyer (1995) has conducted a systematic review of experiments based on the former, including of the Charleston Claimant Placement and Work Test Demonstration, the Washington Alternative Work Search Demonstration, and the New Jersey UI Reemployment Demonstration (cited above), among others. All these interventions seem to work in reducing the duration of UI receipt while providing primarily light-touch services. Moreover, they do not appear to decrease—and may even increase—earnings in the short run, suggesting that workers are not being led to accept lower paying jobs rather than the better jobs to which more protracted job search might have led them. More recent research that aims to target job search assistance on those likely to face long spells of unemployment—such as the study of the Job Search Assistance Demonstration (Decker et al., 2000) and the Worker Profiling and Reemployment Services System (Dickinson et al., 1999; Black et al., 1999)—reach similar conclusions; i.e., that job search assistance can modestly or markedly lower UI duration and benefit receipt and potentially also increase earnings. Finally, extensive work on job assistance services provided to welfare recipients (to be reviewed more carefully below) reinforces these general findings.

These interventions seem to work at least partly—possibly even largely—because the mandatory nature of the intervention provides a hassle factor that some participants seek to avoid; thus, the largest effects on UI exits appear to come quickly, when workers are notified that attendance at a service activity is required, but before they actually receive the service (e.g., Black et al., 1999; Decker et al., 2000). However, researchers have concluded that the services themselves seem important as well (Meyer, 1995; Stanley et al., 1998).

Skill Building

Unfortunately, the evidence on the efficacy of various skill building strategies is not as clear as it might be, due to the absence of very many evaluations that directly tested one training regimen against another in the same setting and with participants of the same type. The JTPA Experiment provides perhaps the most direct evidence, but, even here, researchers tested how effective each training strategy was among *those recommended for* the strategy (relative to a control group also recommended for the same strategy) rather than how effective one strategy would be when compared against another.

Despite these limitations, tentative conclusions can be drawn regarding the efficacy of three common skill-building activities, classroom training, work experience, and on-the-job training. In evaluating the evidence as a whole regarding these three services, Greenberg et al. (2003a) conclude that, at least when considering earnings impacts for adult women, vocational classroom training and OJT are generally effective, work experience is usually effective but less so, and basic skills classroom training is virtually never effective. Results are less clearcut for adult men, youth, and dislocated workers.

In the sections below, we discuss the evidence pertaining to these three service strategies in turn, and follow it with early evidence on the effectiveness of entrepreneurial training, which has only recently been studied systematically.

Classroom Training. The effect of classroom training appears to be quite varied. Basic skills classroom training seems particularly ineffective, as suggested by the Minority Female Single Parent Demonstration (Burghardt et al., 1992), and reinforced by the extensive body of research on welfare-to-work programs (to be reviewed below). More generally, classroom training was found in the JTPA experiment to be among the least effective service strategies (Bloom et al., 1997). On the other hand, when the focus is on job specific skills, classroom training can be very effective, particularly for women and possibly for youth (Greenberg et al., 2003a) or when the instruction provides technical training in fields highly demanded by employers (e.g., Jacobson et al., 2002). A model that also holds some promise is the CET approach of integrated occupational and basic skills training, with intensive training provided in a job-relevant context (Burghardt and Gordon, 1990). However, the effectiveness of this strategy for youth is inconsistent (Miller et al., 2003; Cave et al., 1993).

Work Experience. Work experience, and a related activity, public service employment, have sometimes been disparaged because of their high cost and the presumed absence of clear training content. However, the National Supported Work Demonstration and Homemaker-Home Health Aide Demonstration show that this activity can have large and persistent earnings effects for

welfare recipients (but not for youth), at least when accompanied by a detailed training plan and delivered in a supported environment (Couch, 1992; Bell and Orr, 1994). More generally, Greenberg and colleagues (Greenberg et al., 2003a) have found from their meta-analysis that work experience has a sizable earnings impact in the year after training took place for adult women, but fails to have a significant effect for adult men or youth.

On-the-Job Training (OJT). OJT has for a long while been touted as a particularly effective training strategy. For example, although acknowledging that the evidence was weak, Barnow (1985) concluded in his review of the CETA literature that OJT was among the service activities with the most consistently positive effects for most groups. Results based on the JTPA Experiment also suggest that OJT is among the most efficacious services for both adult women and men for whom this service activity was recommended (Bloom et al., 1997).[13] Greenberg's meta-analysis (Greenberg et al., 2003a) further suggests the effectiveness of this service. On the other hand, Kogan et al. (1991) voice the concern that, in the worst cases, OJT sometimes merely represents a payment to an employer to secure a job spot, and that the training content can be weak. Thus, the mechanism by which OJT exerts its effect is not entirely clear.

Entrepreneurial Training. Self-Employment Assistance (SEA) programs, authorized by Congress for Unemployment Insurance claimants since 1993, provide entrepreneurial training to program participants while waiving the UI program's standard work-search requirements. An experimental evaluation of two demonstration projects (Benus et al., 1994) and a more recent quasi-experimental study of several others (Kosanovich and Fleck, 2001) suggest that SEA can greatly increase the rate of self-employment, but, based on survey data, effects on employment and total earnings (whether from self-employment or not) are uncertain. Although very few UI recipients seem interested in taking up the offer of entrepreneurial training, Benus et al. conclude that SEA programs can nonetheless be a useful policy tool.

Supportive Services

Supportive services are widely assumed to be critically important for promoting the success of a training intervention, but, oddly, its impact has rarely been carefully examined. One exception, though, is the examination of the importance of transportation assistance in the recent Bridges to Work program. Taking note of the often cited "spatial mismatch" between low-skill workers and available jobs (e.g., Wilson, 1987), this project provides transportation services and

[13] The JTPA results regarding OJT are difficult to interpret, because fewer than one-third of those assigned to this service category actually received OJT, and roughly as many received job search assistance.

job search assistance to help inner-city job seekers access jobs in the suburbs, on the premise that providing access to wider employment opportunities would enable them to sustain more attractive employment once their participation in the program ended. However, no effects on earnings were found 18 months after randomization, although program participants were more likely to be employed in jobs that offered health benefits (Roder and Scrivner, 2005). Clearly, though, much more research on the potentially important role of supportive services, as well as case management, is needed.

Lessons from Welfare-to-Work Experiments

We have noted already that the mechanisms by which mandatory programs exert their effect might be very different from those causing voluntary programs to be effective. The difficulty in disentangling the effect of the services actually received as opposed to the effect brought about by those seeking to avoid mandatory participation requirements was already demonstrated with respect to the job search assistance programs targeted to UI claimants, discussed above. A similar difficulty characterizes the large body of research conducted on welfare-to-work programs. Nonetheless, this research has bearing on the issues being discussed in this section, particularly in light of the large number of studies that yield essentially the same conclusion.

Among the most illuminating of these studies for our purposes is the National Evaluation of Welfare-to-Work Strategies (NEWWS), which compared work-first approaches that emphasize providing job search assistance (which the authors call labor-force attachment, or LFA, approaches) with human capital development (or HCD) approaches that largely emphasize basic skills remediation. Head-to-head comparisons of these models in three sites avoid the dangers inherent in making judgments about what service model works best from cross-site comparisons (Hotz et al., 2000) and show that, although both program types were equally effective in reducing welfare, work-first programs produced as large or larger effects on employment and earnings than did the human capital approaches and were much less costly to operate (Hamilton et al., 2001; Hamilton, 2002). This is largely the conclusion reached in a meta-analysis of 27 experimental evaluations of mandatory welfare-to-work programs conducted by Greenberg and colleagues, who find that participation in job search services is positively related to subsequent earnings and negatively related to welfare receipt, while participation in skills training appears to add little (Greenberg et al., 2005b), at least if it involves merely basic skills training (Bloom et al., 2003; Plimpton and Nightingale, 2000; Bloom and Michalopoulos, 2001). The failure of the latter to be efficacious at least in part seems to stem from the fact that few participants in basic skills programs in actuality realized any gains in their basic skills and even fewer earned a credential of any type (Hamilton, 2002).

Although LFA approaches may be effective low-cost strategies, the sobering news is that effects are generally quite small, usually amounting to an earnings boost of no more than about $500 or $600 per year at their peak (e.g., Hamilton, 2002), leading some to conclude that the widespread use of low-cost job search services, coupled with the targeted use of skill-building geared towards a small subset, is the optimal combination (Friedlander and Gueron, 1990).

The Size and Duration of Program Impacts

The above results have suggested that employment and training programs of some types can have significant effects for some groups. But how large do these effects tend to be and how long do they last? The answer to these questions is obviously highly relevant for benefit-cost calculations—participants in programs focused on skill building are likely to experience some foregone earnings during the period while they are in training, but, as the studies reviewed above suggested, may eventually catch-up and surpass their control group counterparts once training ceases. However, unless these earnings gains are sufficiently large and persistent, the short-term boost in earnings that participants may realize potentially will not compensate for the earlier earnings losses and are thus unlikely to make the training investment worthwhile even from the participant's perspective.

The Size of Earnings Effects for Voluntary and Mandatory Programs

We have already seen from Exhibit 1 that estimated annual earnings impacts can range from the negative to positive values of $2,000 or more. Exhibit 2 narrows the focus to just experimental evaluations, to ensure we have the most credible point estimates available. Drawing on this evidence, this table shows, for adult women, adult men, and youth, estimated earnings impacts (expressed in 2003 dollars) in the second year after the year in which random assignment took place, which some have suggested represents the approximate period when earnings effects are near their peak for many programs (Greenberg et al., 2004; Greenberg et al., 2003b; Hamilton, 2002). These are, generally, per enrollee (as opposed to per assignee) estimates, and thus represent the impact on outcomes among those treatment group members who actually obtained some service.[14] Note that a single figure is presented, representing an overall average impact from the study whose results are cited. Most studies report additional impact estimates for various subgroups (e.g., minority vs. white adult women) or for various sites in which the

[14] Not all studies reported results for exactly the third full year after randomization; in these cases, impacts for a part of the third year were used and were annualized. These are all per enrollee impacts, except for NSWD, which is a per assignee impact. If the original source reported a per assignee impact, the per enrollee impact estimate was calculated by dividing the per assignee impact by the program participation rate (the latter was not available for NSWD). Multiple values are reported for some programs (e.g., the JTPA Experiment) because investigators used different data sources (i.e., survey data, Unemployment Insurance wage data, or Social Security earnings records) for measuring outcomes.

experiment was carried out (e.g., the seven sites in which H-HHAD was implemented). The overall average impact was used, if it was reported, or the approximate overall average was computed by taking a weighted average for the separately reported subgroups. As overall averages, though, these figures mask what is sometimes tremendous variation among subgroups or sites; thus, some subgroups or sites recorded impacts much larger or smaller than the averages reported here. However, the average figures shown in the table give a general order of magnitude sense of what types of impacts these groups realized on average.

Consistent with Exhibit 1, the results show earnings impacts that range from the meager, or even negative, to those that are substantial, at a high of about $2,600 per year for the Homemaker-Home Health Aide Demonstration. On average for women, an effect of about $1,300 on a per enrollee basis is about the norm for these studies, with the figure slightly lower for adult men, and with an average effect near zero for youth. Job Corps seems to be the one major bright spot for youth, but, as we have discussed above, these effects quickly dwindle in the subsequent years, except for the group of older youth. Thus, as we have already discussed, adult women seem to benefit the most from publicly-funded employment and training programs, youth the least (or not at all), and adult men somewhere in between.

To put these impact estimates in perspective, Bloom et al. (1997) report that the impacts realized by program participants in JTPA represent about a 16 percent earnings boost for adult women, an 11 percent boost for adult men, and no boost to speak of for youth. Thus, at least for adult women and men, the benefits from participation can be appreciable in percentage terms and could make a notable difference in participants' lives. At the same time, as has often been noted (e.g., Friedlander et al., 1997; LaLonde, 1995), in absolute dollar terms earnings gains of this amount will still generally leave poor families struggling to make ends meet.

The Duration of Effects

Unfortunately, most evaluations of employment and training programs have looked at outcomes for only a year or two after services have ended, so we know little about the extent to which earnings gains persist or the rate at which they decay. However, a thread of work, and more in recent years, is beginning to shed some light on these issues. For example, studies with at least three years of impact data (following the year in which services occurred) include the following:

- Subsequent analyses building on the JTPA experiment tracked impacts for five years (U.S. GAO, 1996) and, later, seven years (Westat, nd) after randomization.
- A number of studies have up to six years of quasi-experimental impact data for MDTA (Ashenfelter, 1978; Bloom, 1984).

Exhibit 2:
Annual Earnings Impacts in the Third Year, for Selected Programs
(Figures in 2003 dollars)

	Program	Primary Service	Avg. Impact
Voluntary Programs for Adult Women			
Bloom et al. (1997)	JTPA	Various	$1,192
USGAO (1996)	JTPA	Various	$1,447
Westat (nd)	JTPA	Various	$932
Burghardt et al. (1992)	MFSPD	Comprehensive	$1,080
Bell and Orr (1984)	H-HHAD	Work exp.	$2,575
Couch (1992)	NSWD	Work exp.	$660
Voluntary Programs for Adult Men			
Bloom et al. (1997)	JTPA	Various	$1,205
USGAO (1996)	JTPA	Various	$1,238
Westat (nd)	JTPA	Various	$493
Voluntary Programs for Youth			
Bloom et al. (1997)	JTPA	Various	-$213
USGAO (1996)	JTPA	Various	-$384
Westat (nd)	JTPA	Various	-$425
Cave et al. (1993)	JOBSTART	Comprehensive	$705
Couch (1992)	NSWD	Work exp.	$443
Schochet et al. (2001)	Job Corps	Comprehensive	$1,182
Schochet et al. (2003)	Job Corps	Comprehensive	$369
Miller et al. (2003)	CET Rep.	Integrated skills	-$695
Quint et al. (1997)	New Chance	Comprehensive	-$118

Note: Acronyms are: MFSPD (Minority Female Single Parent Demonstration), H-HHAD (Homemaker-Home Health Aide Demonstration), NSWD (National Supported Work Demonstration), and CET Rep. (CET Replication Study). See Appendix A for information about these and other programs.

- Social Security data were used to track impacts for eight years after training for women and youth who participated in the National Supported Work Demonstration (Couch, 1992).
- The Minority Female Single Parent Demonstration presents findings for CET through five years (Zambrowski and Gordon, 1993).
- The New Jersey Unemployment Insurance Reemployment Demonstration Project records outcomes for six years (Corson and Haimson, 1995).
- A number of studies provide impact estimates for at least three years for youth programs, including JOBSTART (Cave et al., 1993), LEAP (Bos and Fellerath, 1997), and Job Corps, including from an early quasi-experimental evaluation (Mallar, 1982) as well as from the recent Job Corps experiment (Schochet, et al., 2003), with data for nine years.

Building on findings from these studies and others, Greenberg et al. (2004) developed an estimation model to decipher the trend in annual impacts on earnings based on years elapsed since program enrollment occurred. Their findings are plotted in Exhibit 3.[15]

Given that relatively few evaluations have examined effects for at least three years beyond the year of randomization, and there are so many potentially important variables to consider, firm conclusions are hard to draw. Nonetheless, consistent with earlier discussions in this paper, it seems clear that adult women fare far better than any other group, with annual earnings impacts that are not only much larger than they are for adult men or youth, but with effects that appear to persist for quite a few years after services were received. Indeed, Greenberg et al. (2004) surmise that earnings impacts for adult women may persist indefinitely. By contrast, effects for adult men and youth peak relatively soon after the services were received—in the third year after enrollment for both groups—and then decrease steeply thereafter. At least for adult women, then, publicly-funded employment and training programs can have effects that are quite persistent, offering the prospect of permanently shifting the earnings trajectory. However, no such long-term effects are in evidence for adult men or youth.

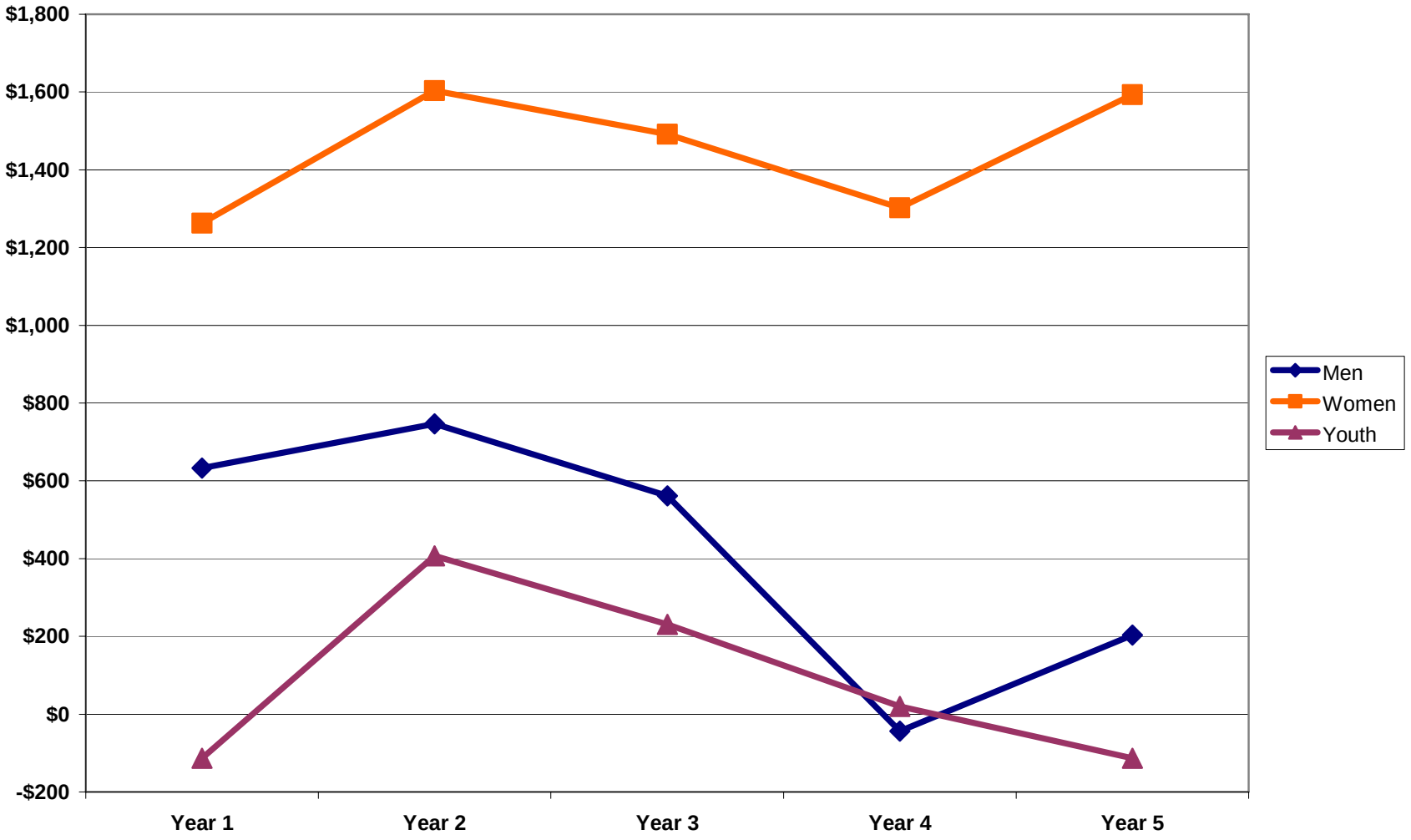
By way of contrast, the earnings impacts for adult women shown in Exhibit 3 can be compared to those derived from recent studies of mandatory welfare-to-work programs for welfare recipients. In perhaps the most recent and comprehensive evidence, Hamilton (2002) shows that annual earnings effects in labor-force attachment programs she studied peak quite quickly after the receipt of services were received. After remaining relatively flat for two more years, the earnings impacts then appear to taper off sharply, declining to near zero after about five years.

These findings are generally quite consistent with those reported by others (e.g., Hotz et al., 2000; Friedlander and Burtless, 1995), including the Greenberg et al. (2003b) meta-analysis of welfare-to-work programs, which concludes that the impacts of mandatory work-first welfare programs peak a couple of years after the intervention and then decline gradually, eventually tapering off to zero after about six years.

---

[15] The X-axis represents the year elapsed following the year in which services began. The chart was developed based on data contained in Greenberg et al.'s (2004) paper, after inflating earnings to 2003 dollars.

Exhibit 3: Estimated Trend in Annual Earnings Impacts



These results suggest, then, that the effects of even light touch job search assistance services can be substantial and surprisingly long-lived. Moreover, some welfare-to-work programs, including the much touted work-first programs in Riverside County and Portland, can realize remarkable success (Hamilton, 2002). When compared with the results in Exhibit 3, however, it appears that, for much larger or more persistent effects for adult women to occur, some skill building will generally need to take place. For example, Couch (1992) shows that the National Supported Work Demonstration, which provided lengthy, closely-supervised work experience, had effects for welfare women that persisted for at least eight years, and that topped $1,000 per year well after services were received. Similarly, as we reported in Exhibit 2, the Homemaker-Home Health Aide Demonstration led to earnings gains that exceeded $2,000 per year (however, the persistence of these latter effects beyond the first few years after services has not been tested).

More generally, Ashenfelter (1978) and Bloom (1984) generated six-year earnings impacts for MDTA that are quite large, and the intensive integrated vocational skills instruction provided by San Jose's Center for Employment and Training has been shown to have sizable long-run effects for different target populations, including minority mothers (Zambrowski and Gordon, 1993) and youth (Cave et al., 1993; but see Miller et al., 2003, for less favorable findings from the CET Replication study). And recent findings from the Job Corps experiment show substantial earnings gains for older youth even nine years after randomization (Schochet and Burghardt, 2005).

Note that certainly not all skill-building programs have effects that persist—indeed, skills training may negatively impact earnings in many instances, and remediation as a stand-alone activity consistently fails to show positive effects. But, under some circumstances, skill-building can yield a sizable and long-lasting earnings boost.

Why Aren't Effects Larger or Longer Lasting?

As noted, even at their peak, impact estimates are generally quite modest and generally experience some decay over time. One might reasonably ask why effects are not larger and why they do not persist indefinitely for all groups. Several explanations might be offered.

First, *many interventions emphasize back-to-work strategies rather than skill-building*. In a recent article, Gottschalk (2005) has presented evidence that work itself can alter individuals' feelings of self-efficacy and, thereby, potentially provide the impetus they need to embark on a productive work career. Similarly, individuals who are helped to establish a toehold in the labor market can thereafter develop the connections and work habits they need to build on their initial success, and may be able to access firm-specific training that improves their earnings prospects

in the long-run. These mechanisms can help explain why work-first programs have effects that last as long as they do.

At the same time, the evidence we reviewed in the preceding section suggests that work-first programs increase earnings only very modestly and typically do not have effects that persist beyond about a half-dozen years (e.g., Greenberg et al., 2003b). Indeed, evidence has shown that interventions focused merely on providing job search assistance are effective because they shorten the duration of unemployment, and thus increase weeks worked, without having much of an effect on the hourly rate of pay (e.g., Plimpton and Nightingale, 2000). Thus, based on this evidence, we should not reasonably expect earnings gains to be very pronounced or to persist indefinitely unless an intervention improves participants' job skills in a fairly substantial way. [16]

Second, *interventions are generally very modest*. Even where skill building is entailed, the interventions are generally quite modest, so the impact on earnings can be expected to be commensurately small. To establish an appropriate yardstick, the returns to an additional year of education are estimated to be in the range of eight to ten percent (LaLonde, 1995; Heckman et al., 1999). However, most employment and training programs provide interventions that are much shorter in duration than a full year of schooling and are much less intensive. For example, Bloom et al. (1997) report that program applicants assigned to the treatment group in the national JTPA experiment received only a few hundred hours of service (267 hours for adult men, up to 438 for female youth) on average, at a per participant cost of only about $2,400. Given this modest investment, Friedlander et al. (1997) conclude that the returns to training in JTPA for both adult women and men were in fact quite large, with the approximately 10 percent to 15 percent earnings boost we previously reported. In the judgement of Heckman et al. (1999), the returns to the job training investment would need to be extraordinarily high for larger effects to emerge and persist. In other words, as LaLonde puts it, "the best summary of the evidence about the impact of past programs is that we got what we paid for" (LaLonde, 1995, pp. 149).

Third, *dropout and substitution effects are often pronounced*. As we discussed early in this paper, experimental evaluations typically proceed by comparing the outcomes of those allowed access to the program being studied relative to those in a control group denied access but who

---

[16] In actuality, it is often quite difficult to disentangle whether an earnings impact comes about because of effects on hours worked rather than the rate of pay, due to the heavy reliance on UI wage records for measuring earnings (which often lack separate measures of weeks worked and rate of pay) and because of the potential selectivity bias to which such an examination might be subjected (e.g., Heckman et al., 1999). For this reason, it is by no means clear whether skill-building programs are any more effective in raising wages than job search programs are. However, see Schochet et al. (2001) for some recent evidence that effects on wages do occur when intensive training takes place.

are free to seek similar services elsewhere. In other words, they examine the effect of being granted access to a program, rather than studying the effect of receiving a service relative to receiving no service at all. In light of this, impact findings can be seriously misleading if used to draw conclusions about the effectiveness of services to the extent that those assigned to the treatment group never in fact receive services or those assigned to the control group access similar services from other sources.

In fact, Heckman and colleagues (Heckman et al. 2000) point out that such "dropout and substitution effects" are commonplace and can be substantial in magnitude. The results we report in Exhibit 4 from the JTPA experiment, which we have chosen to be illustrative, demonstrate this fact. For example, among adult women assigned to the JTPA treatment group, only about 60 percent any employment or training services, while over 30 percent of control group members did so, presumably very often from other publicly funded sources. Thus, the gap in service receipt between treatments and controls was a mere 26 percentage points for this group, and about the same for JTPA adult men and youth.[17]

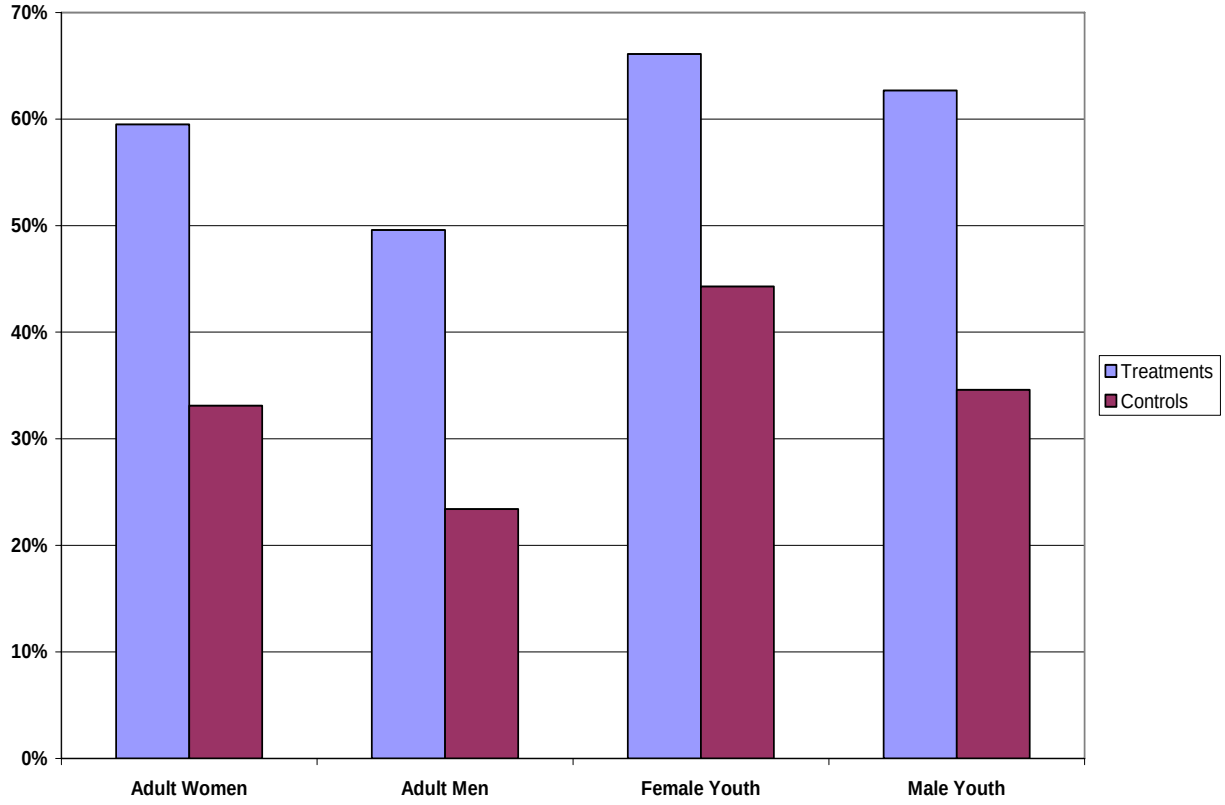
The importance of dropout and substitution effects has been repeatedly demonstrated in evaluations of other programs as well. For example, in their evaluation of New Chance, Quint et al. (1997) remark that the differential in service receipt between treatment and controls was "surprisingly modest." Similarly, in the CET Replication study, more than 52 percent of control group members received some education or employment service sometime in the 30-month follow-up period, compared to 54 percent for those in the treatment group, a non-significant difference of less than 3 percentage points; moreover, the average gap in intensity of participation between the two groups during this period was just 61 hours (Miller et al., 2003).

In the presence of such dropout and substitution effects, the opportunity for treatment group members to record substantially more positive outcomes than their control group counterparts is correspondingly weakened. As Heckman et al. (2000) point out, at the limit, if several programs are close substitutes for each other, it will be very difficult to identify any single program as effective. Moreover, their non-experimental estimates of the JTPA results show that the impact of *receiving training* relative to no training (regardless of whether the individual is in the treatment or comparison group) is well in excess of the experimental estimate of the impact of JTPA per se.

---

[17] The figures plotted in Exhibit 4 are taken from Bloom et al. (1997) and represent the percentage receiving any service from any source.

Exhibit 4: Percent of JTPA Treatments and Controls Receiving Any Service



A fourth reason why earnings effects are not as large as we might hope is that *some participants lack the foundation skills to succeed*. According to economic theory, individuals will generally make optimal investments in education and training on their own, based on what they perceive to be the present value of future benefits. There are some obvious reasons why they might not do so, however, that suggests a role for government intervention. For example, many individuals, especially the economically disadvantaged, may face liquidity constraints that prevent them from consuming as much education and training services as they might ideally like. Similarly, they might experience imperfect information or may be myopic about the value of further skill-building activities.

However, another reason why they might not invest more using their private funds is that they believe the return on their investment will be small. This line of thinking is congruent with the argument put forth by Carneiro and Heckman (2003) that those with weak foundation skills—who often are targeted for public employment and training assistance—will not benefit much

from further training. According to their argument, skill deficits emerge early and, once present, are very difficult to erase. Thus, job training programs might be expected to have attenuated effects for broad categories of those who are targeted for services.[18]

And, finally, as in the private sector, *training effects can be expected to decay under most circumstances*. We have clear evidence that a variety of interventions have effects in the short-run, but are disappointed to learn that the earnings gains might soon begin to decay, such that the earnings effects converge to zero over time, at least for some groups. Our expectation implicitly is that these effects should persist much longer, perhaps indefinitely. However, Lillard and Tan (1992) point out that the effects of private sector training also experience a substantial rate of decay, dwindling to zero within about a dozen years after the training occurred. By this standard, the persistent earnings effects that we observed at least for adult women, as reported in Exhibit 3, are all the more impressive.

What Signs Do We Have Whether Something Might Work?

Given the volatility in impacts from program to program, site to site, and across target groups, it would be useful to know what short-run indicators of performance might reasonably well predict whether a program will in fact turn out to be effective in the longer run. One might hope that programs' official performance measures might be used for this purpose.

Although conclusions have been mixed, on the whole the available evidence is not encouraging. In one of the earliest studies to examine this issue, Gay and Borus (1980) found that termination-based measures of performance bore no relationship whatever to estimated program impacts, and they were only slightly more sanguine about the prospects of using follow-up measures, a conclusion reinforced by Dickinson et al. (1984) in their analysis of the relationship between CETA outcomes and program performance. Geraci and King (1981) and Geraci (1984) dispute these conclusions with their own work that finds, on the contrary, that simple termination based measures of participant outcomes work quite well in predicting program impacts, and that having short-term outcomes measures up to only the third month after program exit adds additional explanatory power. By contrast, more complex change measures or longer-term (up to nine months after exit) outcome indicators appear to add little further.

---

[18] At the same time, there is no real evidence that the so-called hardest-to-serve realize fewer gains from program participation than others, at least insofar as JTPA results are concerned (Heckman et al, 2002; Bloom et al, 1994)

All these efforts were hampered, though, by the fact that program impacts, against which the utility of outcome indicators is being assessed, were developed by using quasi-experimental indicators that are known to measure impacts with substantial error. More convincing evidence, therefore, comes from recent research using the JTPA experimental data, and these results are not at all encouraging. Heckman et al. (2002) find there is no systematic relationship between any of JTPA's performance measures and program impacts, and, in a related effort, Barnow (2000) finds a positive but only very weak relationship. Similarly, in the Job Corps experiment, Burghardt and Schochet (2001) report that a Job Corps center's performance on the program's required performance measures is not related in any systematic way to whether the center is credited with higher or lower program impacts in the several years after youth receive services, nor is it related to impacts in the longer term (Schochet and Burghardt, 2005). In other words, short-term indicators of performance outcomes do not seem to provide any reliable clue as to which programs will be credited with larger program impacts.

However, there are notable deficiencies with even the work that used experimental data, making conclusions tentative. For example, Burghardt and Schochet collapsed centers into a small number of discrete categories, which obscures possible relationships between performance and impacts within each category. And Heckman and Barnow are seriously constrained by the very limited number of observations over which their correlational analyses were conducted, which seriously weakens the statistical power of these methods to detect relationships.

Moreover, even notwithstanding these findings, West (2002) and others (Burghardt and Schochet, 2001; Barnow, 2000) rightly point out that the absence of a clear relationship should not be taken to mean that performance measures are without value, because their very existence might cause all program managers to focus on obtaining satisfactory employment outcomes for their participants in a way that they might not otherwise. In other words, a comparison of what happens to program impacts with and without the application of performance measures has never been properly assessed.

Conclusions

The preceding has provided a broad sweep of a vast body of literature that has attempted to assess the effectiveness of employment and training services for diverse populations. In summary, evidence seems to support the following conclusions:

- Interventions of most types have been found to be effective for economically disadvantaged adult women, and have generally been found to be effective for economically disadvantaged adult men, though with effects that are generally smaller and less persistent than they are for women. On the other hand, youth programs have generally not been successful, even those providing comprehensive services over an extended period of time (though the recent Job Corps findings for older youth represent

an important exception). Evidence regarding the effectiveness of programs for dislocated workers is simply too scanty on which to base firm conclusions.

- Light-touch services, such as job search assistance and job referrals, seem to be remarkably effective in light of their low costs and can have effects that are surprisingly enduring for diverse population groups, including welfare recipients, dislocated workers, and the general population of the unemployed. However, the services have most often been evaluated among populations with a mandatory participation requirement, making the precise mechanism by which effects come about difficult to discern.

- Vocational classroom training, on-the-job training, and work experience have all been shown to be effective in different contexts, though basic skills training—at least for adults as a stand-alone activity—seems to be largely ineffective when judged from the standpoint of its ability to improve subsequent earnings.

- Average annual earnings gains are usually fairly modest, rarely amounting to an earnings boost of more than $1,500, and usually much less—typically substantially less than would be necessary to lift a poor family out of poverty. Nonetheless, effects of this magnitude are impressive in light of the generally modest per-participant costs and short duration of the services that are typically provided.

- When appreciable training is provided, effects for adult women can persist for a substantial period of time, and perhaps indefinitely. However, effects for adult men appear to decay after a few years and dwindle to near zero after about a half-dozen years. Effects for youth are generally near zero to begin with, with the notable exception of those for Job Corps, which appears to have effects for young adults (but not younger youth) that persist for at least nine years, and may even grow over time.

- Dropout and substitution effects are often pronounced, such that evaluations that assess the impact of the offer of services to those in the treatment group will typically seriously understate the actual effect of receiving employment and training services regardless of the source. Thus, evaluation findings are useful in documenting the marginal contribution of the effect of access to the program being studied, but less useful for determining what happens if one undertakes further training or not.

- The effects of training services are highly variable. Many programs, even those providing services that are costly on a per-participant basis, seem to lack efficacy, while others have very large effects (at least relative to the size of the investment) that persist for a half-dozen years and even longer. Moreover, the same program can have vastly different impacts at different sites, presumably depending on features of implementation or the socioeconomic context in which the program is operating. However, we know little about what aspects of implementation matter the most and cannot reliably predict from standard performance indicators which programs are more effective.

This final point underscores the fact that substantial gaps in our knowledge remain, such that, at present, we cannot reliably guide practice to design a program for optimal effect. This

unfortunate circumstance stems from the nature of "black box" evaluations that have used simple measures of an intervention (e.g., whether one is randomly assigned to be allowed access to program services or not), which have not helped us understand very well why something works or not or how it can be improved. This approach may be adequate for programs providing a standard packet of services from site to site, such as Job Corps, which has an established curriculum that is implemented with reasonable fidelity across the nation (Johnson et al., 1999). However, it is much more of a concern when studying formula-funded programs such as WIA and JTPA, which rely heavily on local discretion for service design and delivery. Thus, we know from extensive qualitative research of JTPA and other similar programs that the duration, quality, and characteristics of services that are similar on the surface—such as vocational skills classroom training, work experience, and even job search assistance—can look remarkably different from site to site and represent a very different experience from the customer's perspective (e.g., Kogan et al., 1991).

This extraordinary gap in knowledge is especially problematic given the strong emphasis that WIA places on devolving authority for service design to the local level, on the grounds that local practitioners know best how to meet the needs of their community members. Thus, local areas differ greatly in what types of services count as intensive services under WIA, who gets access to training services, whether training is primarily funded through individual training accounts (ITAs) or contract training, whether providers of training tend to be community colleges or proprietary institutions, and even who gets counted as a WIA registrant (D'Amico and Salzman, 2004a; D'Amico et al., 2004c). Such local discretion is arguably essential for ensuring that local practice best meets local needs and priorities. Unfortunately, though, local workforce boards are making strategic decisions about how to best invest their WIA dollars with little hard evidence to guide them. Perhaps for this reason, Greenberg and colleagues (2003a) see no evidence that employment and training programs have become more effective during the previous three decades.

Implications for Further Research

The above considerations suggest the need for an aggressive research agenda that will help move us away from evaluations that give a summary judgment as to whether an intervention works or not and towards a more fine-grained approach that seeks to understand what components of skill-building activities work best, for whom, and under what circumstances. Some implications of this assessment for further research follow.

1. *Features of the intervention should be empirically linked to variability in program effects.* Knowing whether a complex, variegated program like WIA "works" or not may be helpful from the standpoint of ensuring broad standards of federal

accountability, but arguably is not at all useful for program practitioners attempting to design effective program practices and service strategies at the local level based on the best available research evidence. The reason, of course, is that WIA is not one program but hundreds of separate programs, taking on a distinctly different flavor in each of the nation's local workforce areas. Even more, the intervention can look dramatically different from customer to customer, even within a given workforce area, depending on what services were provided, in what combination or sequence, and with what intensity. To be at all useful on a practical level, our research agenda must acknowledge, and take into account, this diversity and complexity.

Efforts to do so to date have been sparse and of mixed effectiveness. For example, the JTPA experiment made a bold step in this direction by randomly assigning those recommended for each of three service strategies into separate treatment and control groups (Bloom et al., 1994); in other words, random assignment occurred *after* case managers made a service recommendation. However, this effort was only partly successful, because: (a) many of the treatment group members recommended for a service strategy in actuality received something else entirely, a mix of services, or nothing at all, and (b) those who did receive the service for which they were recommended could in actuality have received a treatment of greatly different types, quality, and intensity (e.g., different types of classroom training, in different training fields, and for different durations) than others who were classified as having received the same treatment.

Other efforts show an inkling of the promise that awaits efforts to take a more fine-grained approach. For example, the NEWSS study of welfare recipients was able to document that work-first strategies could be just as effective and cost much less than efforts that focused on basic skills remediation (Hamilton, 2002). Jacobson and colleagues show that the field of study undertaken by dislocated workers enrolled in vocational classroom training matters greatly for the return on investment, with those training for technical fields reaping much greater rewards than others (Jacobson et al., 2002). Similarly, we now have some evidence that integrated basic skills and occupational skills instruction works much better than does teaching these skills in sequence (Burghardt and Gordon, 1990; Cave et al., 1993). In an important methodological innovation associated with the recent Job Corps study, Gritz and Johnson (2001) combine experimental and quasi-experimental methods to document that *completing* assigned project services is critically important for youth to realize any benefits from program participation. Future work that builds on these efforts to provide clues as to *why* an intervention works or does not could substantially inform program practice.

2. Along these lines, *coming to a greater sense of the optimal balance between lighter-touch services and intensive and training services is imperative*. The workforce system under WIA is increasingly coming to rely on making an array of self-services and information tools available as a way of reaching a broad pool of customers in need of publicly-funded workforce services. Although the evidence from voluntary programs is fragmentary, we know from mandatory

programs for welfare recipients (e.g., Hamilton, 2002) and UI claimants (e.g., Meyer, 1995) that such light-touch services can benefit many and are extraordinarily cost effective, though producing earnings effects that are quite modest. These facts imply the need to find an appropriate balance between lighter-touch services and the more costly training services geared to the hard to serve that have long been a hallmark of federal interventions (D'Amico and Salzman, 2004b).

3. *We must come to understand why certain target groups (such as adult women) fare so much better than others (such as adult men or youth) as a consequence of program participation*. Understanding why certain participants benefit more than others is also important for shaping program policy, because it would help us identify the circumstances under which someone could be expected to benefit or not. For example, adult women might benefit more than adult men, and youth might benefit the least, because of the extent of their prior work histories, their opportunity costs of participation, the alternatives they face for seeking equivalent services elsewhere, where they are on their age-earnings trajectory, and whether they are seeking entry-level jobs or careers of one type rather than another. Understanding the role of these various factors would help practitioners target and tailor program services more carefully. For example, knowing why services often *do not* work for youth might also provide clues as to *what would* work.

4. *Program impacts must be measured over a long enough period to enable us to better understand whether earnings effects persist or decay*. So many of our policy conclusions about the effectiveness of interventions are based on evaluations that measure impacts for at best a few years after services were provided. Such findings can be highly misleading, as the sequence of results based on the recent Job Corps evaluation makes very clear (e.g. Schochet et al., 2001; Schochet et al., 2005). Having a longer period of time in which to assess impacts is critical if we are to come to a better understanding of the effectiveness—and the cost-effectiveness—of interventions of various types and for different participant populations.

5. *Continue efforts to identify short-term outcome indicators of longer-term program impacts*. Thus far, we have not been very successful in identifying outcome indicators that relate in any predictable way to estimated program impacts. At the same time, all efforts to identify such relationships have been hampered by severe data limitations. Thus, efforts to identify appropriate short-term outcome indicators should continue, as doing so is imperative for developing a meaningful system of federal performance accountability and giving program managers some guidance on an on-going basis as to whether they are serving their customers effectively or not. In the meantime, there is likely some substantial value in continuing to have some performance indicators in place. However, in recognition of the fact that they have not yet been properly validated, they should be straightforwardly defined, simple to measure, and convey an appropriate message about what the workforce system should value. Arguably, the existing common measures policy achieves these objectives, by laying out a relatively small number of intuitively meaningful indicators that have applicability across a

range of programs, most of which can be measured using existing administrative data and hence which do not require much additional special and costly data collection.[19]

6. *Develop appropriate and reasonable expectations about whether publicly-funded employment and training programs should be judged successful or not.* We have shown in this paper that programs for adult men are often at least moderately successful and that those for adult women are usually quite successful, sometimes even extraordinarily so, given the nature of the barriers to be overcome and the size of the training investments being made. Too often, though, findings like these come up against unreasonable expectations about what short-term and generally low-cost interventions can accomplish (e.g., Mulhausen and Kersey, 2004). Making sure that Congress, the press, and public have reasonable expectations about what can be accomplished will help ensure that the effectiveness of publicly-funded programs is being judged by an appropriate yardstick. Thereafter, funding decisions can be more prudently made, based on whether what can reasonably be accomplished with modest investments is judged to be worth it, in light of competing government priorities.

Overall, we know much, but need to know much more. Expanding on the paucity of the existing evidence in the ways identified above requires a concerted research agenda that draws on theory, takes advantage of natural site-to-site variation in service design and delivery strategies, blends experimental and quasi-experimental methods to draw on the advantages of each, and carefully documents—through implementation studies—the types of interventions that successful customers of the workforce system actually receive. Only in this way can we accumulate a base of knowledge regarding the elements of effective training that can sensibly guide policymakers and practitioners as they go about shaping the next generation of workforce development policies and practices.

[19] For information on the common measures, see *Training and Employment Guidance Letter 17-05*, issued by DOL's Employment and Training Administration on February 17, 2006.

BIBLIOGRAPHY

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics* 60 (Feb): 47-57.

Barnow, Burt. 1987. "The Impact of CETA Programs on Earnings: A Review of the Literature." *Journal of Human Resources* 22 (Spring): 157-193.

Barnow, Burt. 2000. "Exploring the Relationship between Performance Management and Program Impacts." *Journal of Policy Analysis and Management* 19 (Winter): 118-141.

Barnow, Burt S. and Jeffrey A. Smith. 2004. "Performance Management of U.S. Job Training Programs." Pp. 21- 55 in C. O'Leary, R. Straits, and S. Wandner (eds.) *Job Training Policy in the United States*. Kalamazoo, MI: W.E. Upjohn Institute.

Bassi, Laurie J. 1983. "The Effect of CETA on the Postprogram Earnings of Participants." *Journal of Human Resources* 18 (Fall): 539-556.

Bassi, Laurie J. 1984. "Estimating the Effect of Training Programs with Non-random Selection." *Review of Economics and Statistics* 66 (Feb.): 36-43.

Bell, Stephen H. and Larry L. Orr. 1994. "Is Subsidized Employment Cost Effective for Welfare Recipients? Experimental Evidence from Seven State Demonstrations." *Journal of Human Resources* 29 (Winter):42-61.

Benus, Jacob M., Terry R. Johnson, Michelle Wood, Meelima Grover, and Theodore Shen. 1994. *Self-Employment Programs A New Reemployment Strategy: Final Impact Analysis of the Washington and Massachusetts Self-Employment Demonstrations.* Unemployment Insurance Occasional Paper 95-4. Washington, D.C.: U.S. Department of Labor.

Black, Dan A., Jeffrey A. Smith, Mark C. Berger, and Brett J. Noel. 1999. "Is the Threat of Training More Effective than Training Itself? Experimental Evidence from the UI System." University of Western Ontario Department of Economics Working Paper No. 9913.

Bloom, Dan and Charles Michalopoulos. 2001. *How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research*. New York: MDRC.

Bloom, Howard. 1984. "Estimating the Effect of Job-Training Programs, Using Longitudinal Data: Ashenfelter's Findings Reconsidered." *Journal of Human Resources* 14 (Fall): 544-556.

Bloom, Howard, Robert Boruch, Michael Borus, Judith Gueron, Alan Gustman, Peter Rossi, Fritz Scheuren, Marshall Smith, Frank Stafford, and Ernst Stromsdorfer. 1985. *Recommendations of the Job Training Longitudinal Survey Research Advisory Panel*. Washington, D.C.: U.S. Department of Labor.

Bloom, Howard, Larry L. Orr, George Cave, Stephen H. Bell, Fred Doolittle, Winston Lin. 1994. *The National JTPA Study: Overview of Impacts, Benefits, and Costs of Title II-A.* Bethesda, MD: Abt Associates.

Bloom, Howard, Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources* 32 (Summer): 549-576.

Bloom, Howard, Charles Michalopoulos, Carolyn J. Hill, and Ying Lei. 2002. *Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?* New York: MDRC.

Bloom, Howard, Carolyn Hill, and James Riccio. 2003. "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments." *Journal of Policy Analysis and Management* 22 (Fall): 551-575.

Bos, Johannes and Veronica Fellerath. 1997. *LEAP: Final Report on Ohio's Welfare Initiative to Improve School Attendance among Teenage Parents*. New York: MDRC.

Burghardt, John, and Anne Gordon. 1990. *More Jobs and Higher Pay: How an Integrated Program Compares with Traditional Programs*. New York: Rockefeller Foundation.

Burghardt, John, Ana Rangarajan, Anne Gordon, and Ellen Kisker. 1992. *Evaluation of the Minority Female Single Parent Demonstration: Volume I Summary Report*. Princeton, NJ: Mathematica Policy Research.

Burghardt, John, and Peter Schochet. 2001. *National Job Corps Study: Impacts by Center Characteristics*. Princeton, NJ: Mathematica Policy Research.

Carneiro, Pedro and James J. Heckman. 2003. "Human Capital Policy." Pp. 77-239 in Krueger and Heckman (eds.) *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.

Cave, George, Hans Bos, Fred Doolittle, and Cyril Toussaint. 1993. *JOBSTART: Final Report on a Program for School Dropouts*. New York: Manpower Demonstration Research Corporation.

Corson, Walter, David Long, and Walter Nicholson. 1985. *Evaluation of the Charleston Claimant Placement and Work Test Demonstration*. Unemployment Insurance Occasional Paper 85-2. Washington, D.C.: U.S. Department of Labor.

Corson, Walter, Paul T. Decker, Shari Miller Dunstan, and Ann R. Gordon. 1989. *The New Jersey Unemployment Insurance Reemployment Demonstration Project: Final Evaluation Report*. Unemployment Insurance Occasional Paper 89-3. Washington, D.C.: U.S. Department of Labor.

Corson, Walter, Paul Decker, Phillip Gleason, and Walter Nicholson. 1993. *International Trade and Worker Dislocation: Evaluation of the Trade Adjustment Assistance Program*. Princeton, NJ: Mathematica Policy Research.

Corson, Walter, and Joshua Haimson. 1995. *New Jersey Unemployment Insurance Reemployment Demonstration Project: Six Year Follow-up and Summary Report*. Washington, D.C.: U.S. Department of Labor.

Couch, Kenneth. 1992. "New Evidence on the Long-term Effects of Employment Training Programs." *Journal of Labor Economics* 10 (Oct): 380-388.

D'Amico, Ronald and Jeffrey Salzman. 2004a. *An Evaluation of the Individual Training Account and Eligible Training Provider Demonstration: Final Report.* ETA Occasional Paper 2005-02. Washington, D.C.: U.S. Department of Labor.

D'Amico, Ronald and Jeffrey Salzman. 2004b. "Implementation Issues in Delivering Training Services to Adults under WIA". Pp. 101-134 in Christopher J. O'Leary, Robert A. Straits, and Stephen A. Wandner (eds.) *Job Training Policy in the United States*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

D'Amico, Ronald, Gardner Carrick, Kate Dunham, Jennifer Henderson-Frakes, Deborah Kogan, Vinz Koller, Melissa Mack, Micheline Magnotta, Jeffrey Salzman, Dan Weissbein, and Andrew Wiegand. 2004c. *The Workforce Investment Act After Five Years: Results from the National Evaluation of the Implementation of WIA.* ETA Occasional Paper 2004-05. Washington, D.C.: U.S. Department of Labor.

Decker, Paul T., Robert B. Olsen, Lance Freeman, and Daniel H. Klepinger. 2000. *Assisting Unemployment Insurance Claimants: The Long-Term Impacts of the Job Search Assistance Demonstration*. Princeton, NJ: Mathematica Policy Research.

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*. Menlo Park, CA: SRI International.

Dickinson, Katherine P., Terry R. Johnson, and Richard W. West. 1986. "An Analysis of the Impact of CETA Programs on Participants' Earnings." *Journal of Human Resources* 21 (1): 64-91.

Dickinson, Katherine P., Terry R. Johnson, and Richard W. West. 1987. "The Impact of CETA Programs on Components of Participants' Earnings." *Industrial and Labor Relations Review* 40 (April): 430-441.

Dickinson, Katherine P., Paul T. Decker, Suzanne D. Kreutzer, and Richard W. West. 1999. *Evaluation of Worker Profiling and Reemployment Services Systems*. Oakland, CA: Social Policy Research Associates.

Friedlander, Daniel and Judith M. Gueron. 1990. *Are High-Cost Services More Effective than Low-Cost Services? Evidence from Experimental Evaluations of Welfare-to-Work Programs*. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel and Gary Burtless. 1995. *Five Years After: The Long-Term Effects of Welfare-to-Work Programs*. New York: Russell Sage.

Friedlander, Daniel, David Greenberg, and Philip Robins. 1997. "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature* 35 (Dec): 1809-1855.

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

Geraci, Vincent and Christopher King. 1981. *Employment and Training Program Performance: Long-term Earnings Effects and Short-term Indicators.* Austin, TX: Center for the Study of Human Resources, Univ. of Texas.

Geraci, Vincent. 1984. *Short-term Indicators of Job Training Program Effects on Long-term Participant Earnings.* Austin, TX: Center for the Study of Human Resources, Univ. of Texas.

Glazerman, Steven, Dan M. Levy, and David Myers. 2002. "Nonexperimental Replications of Social Experiments: A Systematic Review." Washington, D.C.: Mathematica Policy Research, Inc.

Gottschalk, Peter. 2005. "Can Work Alter Welfare Recipients' Beliefs?" *Journal of Policy Analysis and Management* 24 (Summer): 485-498.

Greenberg, David, Charles Michalopoulos, and Philip Robins. 2003a. "A Meta-Analysis of Government-Sponsored Training Programs." *Industrial and Labor Relations Review* 57 (Oct): 31-53.

Greenberg, David, Karl Ashworth, Andreas Cebulla, and Robert Walker. 2003b. "Do Welfare-to-Work Programmes Work for Long?" *Fiscal Studies* 25 (Feb): 27-53.

Greenberg, David, Charles Michalopoulos, and Philip Robins. 2004. "What Happens to the Effects of Government-Funded Training Programs Over Time?" *Journal of Human Resources* 34 (1): 277-293.

Greenberg, David H., Charles Michalopoulos, and Philip K. Robins. 2005a. "Do Experimental and Nonexperimental Evaluations Give Different Answers about the Effectiveness of Government-Funded Training Programs?" Draft.

Greenberg, David, Andreas Cebulla, and Stacey Bouchet. 2005b. *Report on a Meta-Analysis of Welfare-to-Work Programs*. Draft.

Gritz, Mark and Terry Johnson. 2001. *National Job Corps Study: Assessing Program Effects on Earnings for Students Achieving Key Program Milestones*. Seattle: Battelle Memorial Institute.

Grossman, Jean Baldwin and Cynthia Sipe. 1992. *Report on Long-Term Impacts*. Philadelphia: Public/Private Ventures.

Grubb, W. Norton. 1996. *Learning to Work: The Case for Reintegrating Job Training and Education*. New York: Russell Sage.

Hamilton, Gayle and Daniel Friedlander. 1989. *Saturation Work Initiative Model in San Diego: Final Report*. New York: MDRC.

Hamilton, Gayle, Stephen Freedman, Lisa Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adams-Ciardullo, Anna Gassman-Pines, Sharon McGroder, Martha Zaslow, Surjeet Ahluwalia, and Jennifer Brooks. 2001. *National Evaluation of Welfare-to-Work Strategies: How Effective are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs.* New York: MDRC.

Hamilton, Gayle. 2002. *Moving People from Welfare to Work: Lessons from the National Evaluation of Welfare-to-Work Strategies*. New York: MDRC.

Heckman, James J. and Jeffrey Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9 (Spring): 85-110.

Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 64 (4): 605-654.

Heckman, James J., Robert J. LaLonde, and Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." Pp. 1865-2095 in Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics Vol. 3A*. New York: Elsevier.

Heckman, James, Neil Hohmann, and Jeffrey Smith. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *Quarterly Journal of Economics* (May): 651-694.

Heckman, James J., Carolyn Heinrich, and Jeffrey Smith. 2002. "The Performance of Performance Standards." *Journal of Human Resources* 37 (4): 778-811.

Hollister, Robinson, Peter Kemper, and Rebecca Maynard (eds.). 1984. *The National Supported Work Demonstration.* Madison, WI: Univ of Wisconsin Press.

Hollister, Robinson. 1990. *The Minority Female Single Parent Demonstration: New Evidence about Effective Training Strategies*. New York: Rockefeller Foundation.

Hotz, V. Joseph, Guido W. Imbens, and Jacob A. Klerman. 2000. *The Long-Term Gains from GAIN: A Re-analysis of the Impacts of the California GAIN Program.* National Bureau of Economic Research Working Paper 8007.

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review* (83 (Sept.): 685-709.

Jacobson, Louis and Ian Petta. 2000. *Measuring the Effect of Public Labor Exchange (PLX) Referrals and Placements in Washington and Oregon: Final Report*. Rockville, MD: Westat.

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 2002. "Measures of Program Performance and the Training Choices of Displaced Workers." Pp. 187-213 in R. Eberts, C. O'Leary, and S. Wandner (eds.) *Targeting Employment Services*. Kalamazoo, MI: W.E. Upjohn Institute

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 2003. *Should We Teach Old Dogs New Tricks? The Impact of Community College Retraining on Older Displaced Workers*. Federal Reserve Bank of Chicago Working Paper 2003-25.

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 2005. "Is Retraining Displaced Workers a Good Investment?" 2005. *Federal Reserve Bank of Chicago Economic Perspectives* (2nd Quarter).

Johnson, Terry R., Katherine P. Dickinson, and Richard W. West. 1985. "An Evaluation of the Impact of ES Job Referrals on Applicant Earnings." *Journal of Human Resources* 20 (Winter): 117-137.

Johnson, Terry, Mark Gritz, Russell Jackson, John Burghardt, Carol Boussy, Jan Leonard, and Carlyn Orians. 1999. *National Job Corps Study: Report on the Process Analysis.* Seattle: Battelle Human Affairs Research Centers.

King, Christopher T. 2004. "The Effectiveness of Publicly Financed Training in the United States: Implications for WIA and Related Programs." Pp. 57-99 in Christopher J. O'Leary, Robert A. Straits, and Stephen A. Wandner (eds.) *Job Training Policy in the United States*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Kodrzycki, Yolanda. 1997. "Training Programs for Displaced Workers: What Do They Accomplish?" *New England Economic Review* (May/June): 39-59.

Kogan, Deborah, Katherine Dickinson, Barbara Means, and Marlene Strong. 1991. *Improving the Quality of Training Under JTPA.* U.S. Department of Labor Employment and Training Administration Research and Evaluation Report Series 91-A.

Kosanovich, William T. and Heather Fleck. 2001. *Final Report: Comprehensive Assessment of Self-Employment Assistance Programs.* Arlington, VA: DTI Associates, Inc.

Krueger, Alan B. 2003. "Inequality, Too Much of a Good Thing." Pp 1-75 in Krueger and Heckman (eds.) *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.

LaLonde, Robert J. 1995. "The Promise of Public Sector-Sponsored Training Programs." *Journal of Economic Perspectives* 9 (Spring): 149-168.

Leigh, Duane E. 1990. *Does Training Work for Displaced Workers?: A Survey of Existing Evidence.* Kalamazoo, MI: W.E. Upjohn Institute.

Leigh, Duane E. 2000. "Training Programs for Dislocated Workers." Pp. 227-260 in B. Barnow and C. King (eds.) *Improving the Odds: Increasing the Effectiveness of Publicly Funded Training*. Washington, D.C.: Urban Institute Press.

Lerman, Robert. 2000. "Employment and Training Programs for Out-of-School Youth: Past Effects and Lessons." Pp. 185-208 in B. Barnow and C. King (eds.) *Improving the Odds: Increasing the Effectiveness of Publicly Funded Training*. Washington, D.C.: Urban Institute Press.

Lerman, Robert, Signe-Mary McKernan, and Stephanie Riegg. 2004. "The Scope of Employer-Provided Training in the United States: Who, What, Where, and How Much?" Pp. 211-243 in Christopher J. O'Leary, Robert A. Straits, and Stephen A. Wandner (eds.) *Job Training Policy in the United States*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Lillard, Lee and Hong Tan. 1992. "Private Sector Training: Who Gets It and What are Its Effects." *Research in Labor Economics* 13: 1-62.

Mallar, Charles, Stuart Kerachsky, Craig Thornton, and David Long. 1982. *Evaluation of the Economic Impact of the Job Corps Program*. Princeton, NJ: Mathematica Policy Research.

McClanahan, Wendy S., Cynthia L. Sipe, and Thomas J. Smith. 2004. *Enriching Summer Work: An Evaluation of the Summer Career Exploration Program*. Philadelphia: Public/Private Ventures.

Meyer, Bruce D. 1995. "Lessons from the U.S. Unemployment Insurance Experiments." *Journal of Economic Literature* 33 (March): 91-131

Miller, Cynthia, Johannes M. Bos, Kristin E. Porter, Fannie M. Tseng, Fred C. Doolittle, Deana N. Tanguay, and Mary P. Vencill. 2003. *Thirty Month Findings from the Evaluation of the Center for Employment Training Replication Sites*. New York: MDRC.

Mulhausen, David and Paul Kersey. 2004. *"In the Dark on Job Training: Federal Job-Training Programs Have a Record of Failure."* Heritage Foundation Backgrounder No. 1774.

Nudelman, Jodi. 2000. "The Impact of Job Training Partnership Programs for Adult Welfare Recipients." Pp. 101- 128 in B. Barnow and C. King (eds.) *Improving the Odds: Increasing the Effectiveness of Publicly Funded Training*. Washington, D.C.: Urban Institute Press.

Organization for Economic Cooperation and Development. 2001. *OECD Employment Outlook 2001: Reconciling Social and Employment Goals*. Paris: Organization for Economic Cooperation and Development.

O'Leary, Christopher J., Paul T. Decker, and Stephen A. Wandner. 2002. "Targeting Reemployment Bonuses." Pp. 161-182 in R. Eberts, C. O'Leary, and S. Wandner (eds.) *Targeting Employment Services*. Kalamazoo, MI: W.E. Upjohn Institute.

O'Leary, Christopher. 2004. "Evaluating the Effectiveness of Labor Exchange Services." Pp. 135-177 in D. Balducchi, R. Eberts, and C. O'Leary (eds.), *Labor Exchange Policy in the United States*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research

O'Neill, Dave M. and June Ellenoff O'Neill. 1997. *Lessons for Welfare Reform*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research

Plimpton, Lisa and Demetra Smith Nightingale. 2000. "Welfare Employment Programs: Impacts and Cost-Effectiveness of Employment and Training Activities." Pp. 49-100 in B. Barnow and C. King (eds.) *Improving the Odds: Increasing the Effectiveness of Publicly Funded Training*. Washington, D.C.: Urban Institute Press.

Polit, Denise, Janet Kahn, and David Stevens. 1985. *Final Impacts from Project Redirection: A Program for Pregnant and Parenting Teens.* New York: MDRC.

Quint, Janet, Johannes Bos, Denise Polit. 1997. *New Chance: Final Report on a Comprehensive Program for Young Mothers in Poverty and their Children*. New York: MDRC.

Riccio, James, Daniel Friedlander, and Stephen Freedman. 1994. *GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program*. New York: Manpower Demonstration Research Corporation.

Roder, Anne and Scott Scrivner. 2005. *Seeking a Sustainable Journey to Work: Findings from the National Bridges to Work Demonstration*. Philadelphia: Public/Private Ventures.

Schirm, Allen and Nuria Rodriquez-Planas. 2004. *The Quantum Opportunity Program Demonstration: Initial Post-Intervention Impacts*. Princeton, NJ: Mathematica Policy Research.

Schochet, Peter Z., John Burghardt, and Steven Glazerman. 2001. *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes*. Princeton, NJ: Mathematica Policy Research.

Schochet, Peter Z., Sheena McConnell, and John Burghardt. 2003. *National Job Corps Study: Findings Using Administrative Earnings Records Data*. Princeton, NJ: Mathematica Policy Research.

Schochet, Peter Z. and John Burghardt. 2005. *Conducting a Long-Term Follow-Up of National Job Corps Study Samples: Impact Findings Using 1993 to 2003 Summary Earnings Records Data*. Princeton, NJ: Mathematica Policy Research.

Social Policy Research Associates. 2005. *Evaluation of the Trade Adjustment Assistance Program: Design Report.* Oakland, CA: Social Policy Research Associates and Mathematica Policy Research, Inc.

Stanley, Marcus, Lawrence Katz, and Alan Krueger. 1998. *Developing Skills: What We Know About the Impacts of American Employment and Training Programs on Employment, Earnings, and Educational Outcomes.* Report for the G8 Economic Summit.

U.S. Department of Labor. 2005. *Workforce System Results*. Available at http://www.doleta.gov/performance/results/Reports.cfm

U.S. General Accounting Office. 1996. *Job Training Partnership Act: Long-Term Employment and Earnings Outcomes*. GAO/HEHS-96-40.

West, Richard. 2002. Comments at the W.E. Upjohn Conference on Job Training Policy in the United States. Augusta, MI.

Westat. nd. *JTPA National Evaluation: Seven Year Follow-up*.

Wilson, William Julius. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago: Univ of Chicago Press.

Zambrowski, Amy, and Anne Gordon. 1993. *Evaluation of the Minority Female Single Parent Demonstration: Fifth-Year Impacts at CET*. Princeton, NJ: Mathematica Policy Research.

Appendix A:
Summary of Key Program Features

Model	Service	Relevant Report	Study Intake Period	Description	Status	Target Population	Sites in Which Evaluated	Funding Source
Bridges to Work	Transportation assistance, job search assistance	Roder & Scrivner (2005)	1997-1999	Provides transportation assistance and job search assistance to help inner-city residents obtain employment in the suburbs, and hence address the supposed "spatial mismatch."	Voluntary	Inner-city job seekers	4 metro areas	HUD & various foundations (Ford, MacArther, Rockefeller)
CET Replication	Integrated basic and occupational skills instruction	Miller et al. (2003)	1995-1999	Provides integrated basic and occupational skills instruction in a worklike setting, open-entry and open-exit, requires intensive participation until competencies are learned.	Voluntary	Disadvantaged out-of-school youth (ages 16-22)	12 sites	DOL
Comprehensive Employment and Training Act (CETA)	Classroom training, OJT, work experience and other services	Bassi (1983) Dickinson et al. (1986) Barnow (1987)	1975-1977 (varies by study)	A national program enacted in 1973 that provided a range of services matched to each participant's needs.	Voluntary	Economically disadvantaged adult women, men, and youth	National	DOL
GAIN (Greater Avenues for Independence, operating as CA's JOBS program).	W2W (JOBS)	Riccio et al. (1994)	1988-1990	Those who are basic skills deficient undergo remediation, those who are skills proficient undergo job search assistance. Those still unemployed after their primary activity undergo further services based on an assessment of need.	Mandatory	Welfare women & UPs	6 CA counties	State of California
Homemaker-Home Health Aide Demonstration	Training and supported work	Bell & Orr (1994)	1983-1986	Several weeks of CRT in home health care, with an on-site practicum, followed by up to 12 months of subsidized work experience.	Voluntary	AFDC women	7 states	HHS

Appendix A:
Summary of Key Program Features

Model	Service	Relevant Report	Study Intake Period	Description	Status	Target Population	Sites in Which Evaluated	Funding Source
Job Assistance Demonstration	Job search assistance	Decker et al. (2000)	1995-1996	Provided three different regimens to targeted UI claimants: job search assistance, individualized job search assistance, and individualized job search assistance with the offer of training.	Mandatory	Unemployed	A number of offices in D.C. and FL	DOL
JOBSTART	Comprehen-sive	Cave et al. (1993)	1985-1988	Modeled on Job Corps, but without a residential component. Included basic skills instruction, occ skills training, support services, and job placement assistance.	Voluntary	Low income dropouts ages 17-21	13 sites	DOL, with additional foundation & corporate support
Job Training Partnership Act (JTPA)	Classroom training, OJT, work experience and other services	Bloom et al. (1997) US GAO (1996)	1987-1989	A national program enacted in 1982 that provided a range of services matched to each participant's needs.	Voluntary	Economically disadvantaged adult women, men, and youth	16 local areas	DOL
LEAP (Ohio's Learning, Earning, and Parenting Program)	Financial incentives	Bos and Fellerath (1997)	1990-1991	Provided financial incentives to teens who stay in school or, if they have dropped out, to return to school, and applies sanctions (in the form of reduced welfare payments) to those who fail to do so.	Mandatory	Teen parents on welfare	12 counties in Ohio	State of Ohio, with additional funding from HHS and various foundations
Manpower Development and Training Act (MDTA)	Training	Ashenfelter (1978) Bloom (1984)	1964	A national program begun in 1962 that provided retraining to the unemployed and disadvantaged.	Voluntary	Unemployed and disadvant-aged	National	DOL
Minority Female Single Parent Demonstration	Comprehen-sive	Burghardt et al. (1992)	1984-1987	Included basic skills instruction, job skills training, job-placement assistance, counseling, child care assistance.	Voluntary	Minority single mothers	4 cities	Rockefeller Foundation

Appendix A:
Summary of Key Program Features

Model	Service	Relevant Report	Study Intake Period	Description	Status	Target Population	Sites in Which Evaluated	Funding Source
National Supported Work Demonstration	Supported work	Hollister et al. (1984) Couch (1992)	1975-1979	Provided supported work, with a gradually increasing semblance to an actual workplace, for up to 18 months.	Voluntary	AFDC women, ex-offenders and ex-addicts, and youth	10 sites	Six federal agencies & the Ford Foundation
New Chance	Comprehen-sive	Quint et al. (1997)	1989-1992	Included basic skills instruction, career exposure, and health and family planning (Phase I), and occ skills training, work experience, and job placement assistance (Phase II), in a sequence lasting up to 18 months.	Voluntary	Teen mothers who were dropouts and on AFDC	16 sites in 10 states	DOL, with additional foundation and corporate support
New Jersey Reemployment Demonstration	Job search assistance, plus training to some	Corson et al. (1989)	1986-1987	An experimental design with three treatment groups: (1) job search assistance alone, (2) job search assistance plus the offer of training, and (3) job search assistance plus a reemployment bonus.	Mandatory	Dislocated workers	State of NJ	DOL & State of New Jersey
Project Redirection	Comprehen-sive	Polit et al. (1985)	1980-1983	Offered educational, employment-related, health, recreational, family planning, parenting, and life-skills components. Strong role for mentorship and peer group support. Also provided financial incentives. Youth can participate for up to 18 months.	Voluntary	Teen mothers without a hs diploma and on AFDC, including some still in school	4 cities	DOL, National WIN Office, and the Ford and W.T. Grant Foundations
Quantum Opportunity Program	Comprehen-sive after-school program	Shirm & Rodriquez-Planas (2004)	1995-2001	Predominately an after-school program providing case management, mentoring, remediation, developmental activities, community service, and financial incentives for up to five years.	Voluntary	At-risk teens entering high-school with low grades	7 sites	DOL and the Ford Foundation

Appendix A:
Summary of Key Program Features

Model	Service	Relevant Report	Study Intake Period	Description	Status	Target Population	Sites in Which Evaluated	Funding Source
STEP (Summer Training & Education Program)	Summer work experience, remediation	Grossman & Sipe (1992)	mid-1980s	Part-time summer work experience and remediation for 2 summers, plus some support during the school year.	Voluntary	Low-income 14-15 yr. olds	5 cities	DOL, HHS, & various foundation
Summer Career Exploration Program	Summer work experience	McClanahan et al. (2004)	1998	Summer work experience in private-sector jobs.	Voluntary	Low-income in-school youth	Philadelphia	William Penn Foundation, with other foundation & corporate support

Appendix B:
Annotations

Report	Program	Service	Target	Results
Ashenfelter (1978)	MDTA	Classroom training	Adults	1) For females, effects of between $300 to $600 (in 1965 dollars) in the year just after training, and these effects persist through the 5-yr period. 2) For males, effects of between $150 to $500 in the year just after training, decreasing to half that amount after 5 years.
Barnow (1987)	CETA	Varied	Mixed	Great variation in CETA impact estimates shows how sensitive findings are to the methods used.
Barnow (2000)	JTPA	Varied	Mixed	1) The relationship between JTPA 10-month regression-adjusted performance and 10-month impacts for earnings measured for adults is modestly positive. The simple correlation for the 16 SDAs is .357 at 10 months, which is not significant. The relationship is weaker for earnings among welfare recipients, and hours worked for all adults, but modest for hours worked among welfare adults. 2) Also examined the relationship between 10-month measured performance and impacts measured longer term (in months 19-30). Results suggest that short-term (i.e., 10-month) measured performance is a poor predictor of longer-term impacts. 3) Although the relationship between performance and impacts is weak, the author recommends against abandoning performance measures as worthless. This is because performance measures could cause SDAs to focus on performance in a way that they wouldn't in the absence of them.
Barnow and Smith (2004)	Mixed	NA	NA	Seven studies have looked at the efficiency implications of performance measures by estimating the correlation between them and program impacts. Findings are mixed or negative—commonly used performance measures do not improve program efficiency by inducing service to those who will benefit most. Further, creaming has neither much of an efficiency benefit or efficiency loss.
Bassi (1983)	CETA	Varied	Mixed	1) CETA has positive and often significant effects on earnings of participants. These effects are usu larger for women than men (and usu are not significant for men). Possibly this is because CETA prepares people for entry-level jobs and this doesn't help men much. 2) No activity seems more effective than any other. However, PSE is extremely costly so is probably not worth it. 3) The disadvantaged may benefit less than the less disadvantaged, but it is hard to tell 4) We really need data for longer-term impacts to adequately judge these programs.
Bassi (1984)	CETA	Varied	Mixed	This paper presents results for white and minority women and minority men. Shows significant effects for women, larger for whites than nonwhites, but no significant effects for minority men.
Bell and Orr (1994)	Homemaker-Home Health Aide Demo	Training and subsidized employment	AFDC adults	Significant earnings gains in the second year after treatment (year 3) in 5 of the 7 states, amounting to $100 to over $200 per month ($1,200 to nearly $2,600 per year in 1984 dollars). Impacts were greater in year 3 than year 2 by a considerable margin.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Black et al. (1999)	WPRS	Job search assistance	Unemployed	WPRS reduces mean weeks of UI benefit receipt by about 2.2 weeks and UI benefits received by about $143, and increases earnings by over $1000. Much, but not all, of the effect results from an increase in early exits from UI coinciding with when claimants found out about their mandatory program obligations.
Bloom & Michalopoulos (2001)	W2W	Varied	Welfare women	1) Programs providing only mandatory employment services were effective, but the most successful used a mix of services, including some education and training. All emphasized the need to find work. Thus, an individualized approach may be best. Programs that included only mandatory employment services left families no better off financially. Most of the increases in earnings were the result of increases in employment (hourly wages didn't change much). That education-focused programs weren't much more effective may be due to the fact that they focused on basic skills rather than voc skills and that few participants actually gained much in the way of skills. 2) Only programs that provided earnings supplements (e.g., earnings disregards) left families better off financially. 3) Relatively little is know about the effects of time limits, but preliminary evidence suggests they do not lead to widespread hardship.
Bloom (1984)	MDTA	Classroom training	Adults	Respecifying Ashenfelter's (1978) model shows much more positive effects for men than Ashenfelter. Effects are larger than Ashenfelter found (for both men and women) and there is no sign of decay for either.
Bloom et al (1997)	JTPA	Varied	Mixed	1) For adult women, the total impact per assignee over the 30-month follow up period was $1,176, a 9.6% increase. Per enrollee, the impact is $1,837. For adult women, differences are small in the in-program period, rise to about $825 in year one, and remain there for year two. 2) Adult men experienced smaller (but still significant) impacts, of $978 per assigned (5.3%), or $1,599 per enrollee. For adult men, differences are small in the in-program period, rise to $500 in year 1, and rise still further to $856 in year two. 3) Differences for female youth and male youth non-arrestees were near zero (and negative for male youth). For female youth, impacts rise over time but remain small even by year 2; for male youth, impacts are increasingly negative. 4) Impacts by treatment strategy vary. For adult women, Other Services work best ($3,949 over 30 months), followed by OJT/JSA ($2,292). CRT doesn't appear to work ($630). For adult men, no strategy is statistically significant. However, OJT/JSA works best ($2,109), followed by CRT ($1,287) and Other Services ($941). 5) Among other noteworthy subgroup impacts, AFDC women benefit appreciably, especially from OJT/JSA. 6) These findings are generally in line with other studies--modest impacts for adult women and men, no effects for out-of-school youth (e.g., JOBSTART, NSWD).

Appendix B:
Annotations

Report	Program	Service	Target	Results
Bloom, Hill, and Riccio (2003)	W2W	Varied	Welfare women	1) Conducts a meta-analysis to examine the effect of management practices, client activities, labor market conditions, and client characteristics on program impacts. Outcomes are measured for the first 2 years after random assignment. 2) With regard to management practices, shows that staffs' emphasis on getting clients a job quickly is extremely powerful. 3) Among other program practices, personalized attention is important, as is smaller caseloads. 4) With respect to services, whether the individual (control or treatment group member) accesses job search services is not particularly important, but receiving basic education services depresses earnings. 5) Impacts are lower when the unemployment rate is higher, suggesting jobs must be available. 6) Some client characteristics matter, but not in a predictable way that would enable one to assert that impacts are lower/higher for the harder-to-serve. In other words, clients of many different types benefit about equally.
Bos and Fellerath (1997)	LEAP	Financial incentives	Teen parents on welfare	Financial incentives and sanctions significantly increased school attendance and the rate at which teens advanced from one grade to the next. Longer-term impacts were less favorable. Teens already in school at program enrollment increased their receipt of GEDs (but not high school diplomas) and realized an earnings boost, but only for the first two (of the four) years studied. There were no impacts for those who were out of school at enrollment on either high school graduation, GED attainment, or post-program employment outcomes.
Burghardt & Gordon (1990)	MFSPD	Varied	Minority single mothers	Reports results for CET vs. the other Minority Female Single Parent Demo sites and notes that CET performs much better, suggesting the importance of integrated skills instruction.
Burghardt & Schochet (2001)	Job Corps	Varied	Youth	1) Impacts were similar across centers of all types. Of special relevance, impacts were similar for centers with high, medium and low performance, based on JC performance measurement system. Outcomes were better at the high-performance centers (by definition), but so too were the outcomes of control group members who would have attended the high-performance centers. 2) We cannot conclude that performance measures are irrelevant, though, because the focus on performance could improve the quality of services for all students.
Burghardt et al (1992)	MFSPD	Varied	Minority single mothers	MFSPD operated at 4 sites. Only one site, CET in San Jose, had significant positive effects on earnings, which were in evidence up to 2.5 yrs after application (though there were no effects on reduced welfare receipt). The remaining three programs produced no significant impacts on earnings. Reasons for CET's greater success may be because it relied on integrated basic and occ skills instruction, rather than the sequential training (basic skills training first, followed by job skills training) favored by the other programs, esp. since nearly half of participants in these programs never made it to job skills training. The findings suggest that moving poor single mothers into the classroom to learn basic skills might not be fruitful.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Carneiro & Heckman (2003)	Varied	Varied	Mixed	1) The return to a year of schooling exceeds 10%. This return is higher for more able people and for children from better backgrounds. 2) Small-scale studies, such as Perry Preschool, show that intensive early investments in young children from disadvantaged backgrounds show remarkable success. Less intensive programs, such as Head Start, also hold promise, but many effects appear to dissipate. 3) Even though cognitive skills are fairly well set by age 8, some interventions in adolescent years can be effective as well. QOP, LEAP, and TPD can all help keep kids from dropping out. So too can financial incentives. On the other hand, dropout recovery programs don't appear successful 4) Job training for older workers and displaced workers don't appear successful. On the other hand, CRT for adults appears to have substantial returns. 5) You get what you pay for. JTPA impacts are small, but so too is the investment 6) Deficits in skills emerge early and are tough to erase. Thus, efforts should be placed on family policy.
Cave et al. (1993)	JOBSTART	Varied	Disadvantaged youth	1) Targets 17-21 year old economically disadvantaged dropouts in 13 sites. Services are modeled on Job Corps but are less intensive and are non-residential. 2) Overall, the program increased the rate of GED attainment, but had no effects on earnings overall--earnings were lower for participants in year 1, and overtook those in the control group in years 2 and 3 but were not significant in those years. Definitely not cost effective, in light of the lost wages in year 1 and modest wage boost thereafter. 3) One program, CET, is an exception and had large earnings gains in year 3.
Corson et al. (1989)	NJ Reemployment Demonstration	Job search and training	Dislocated workers	This experimental design with three treatment groups shows that all three intervention types succeed in reducing UI receipt and increasing employment and earnings in the year following the initial UI claim. However, job search assistance alone worked as well, and at lower cost, than job search assistance coupled with the offer of job training, though its effects seem to decay quickly. Job search assistance alone worked best for dislocated workers with readily marketable skills, but not as well for dislocated workers who lacked marketable skills due to structural shifts in the labor market. For the later group, more intensive services may be necessary.
Corson and Haimson (1995)	NJ Reemployment Demonstration	Job search and training	Dislocated workers	This follow-up study to Corson et al., (1989) measured earnings six years after the initial UI claim, thus allowing a much longer time to assess impacts. All three treatment groups showed earnings impacts in year six. Overall, the JSA plus training group showed no greater earnings gains than those who received JSA alone. However, this can be attributed to the relatively small number of claimants who actually took up the training offer. Supplementary analysis suggests that training--both classroom training and OJT--did enhance trainees' earnings.
Corson et al. (1993)	Trade Adjustment Assistance	Retraining	Dislocated workers	Finds that trainees have higher employment rates and earnings at the end of 12 quarters than a comparison group of non-trainees. However, this difference is attributed to differences in the observable characteristics of the two groups; once controlling for them, no earnings effect remains.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Couch (1992)	National Supported Work Demonstration	Subsidized employment	AFDC adults, and Disadvantaged Youth	1) Results for AFDC Adults: -- Consistent significant and positive impacts, on the order of $250 to $450 per year (in 1978 dollars) -- Impact seems to peak about 5 yrs; the authors conclude that there is little decay -- Based on these results, the cost-benefit calculation is even more favorable than initially thought (initial estimates assumed some rate of decay). 2) Results for Youth: -- No significant effects in any year and sign is not consistently positive.
Decker et al. (2000)	Job Search Assistance Demonstration	Job search assistance	Unemployed	Offered three service strategies: structured job search assistance, individualized job search assistance, and individualized job search assistance with the offer of training. Participation in activities that were not mandatory, such as job search workshops and training, were quite low. All treatments groups recorded lower UI benefit durations in the initial benefit year, but not subsequently. Impacts on employment and earnings were uneven. There was no evidence that the treatments pushed claimants into lower-quality jobs. Effects on UI exit rates occurred early in the benefit spell, suggesting much of the effect derives from the participation requirement itself rather than from skills learned during program participation.
Dickinson, et al. (1986)	CETA	Varied	Mixed	Results show that CETA is 1) Not beneficial for adult men, and may even lower earnings (-$690). All services have negative coefficients. 2) Adult women earn a non-significant $13 more. With respect to services, PSE is beneficial for women, while work experience is negatively related to outcomes for them. Other services are not significant (CRT, OJT, referral). However, different specifications of the matching procedure yield higher and significant impact estimates, on the order of $500.
Dickinson, et al. (1987)	CETA	Varied	Mixed	1) For adult women, CETA has an overall positive effect of $1320. A major factor accounting for this is their increased employment, while adult men are less likely to be employed. 2) Among those employed, for adult women, CETA is associated with a large positive impact on hours worked per week and weeks worked (but no effect on hourly wage). Among employed men, CETA appears to depress weeks worked and hourly wages.
Dickinson et al. (1999)	WPRS	Job search assistance	Unemployed	Impacts of WPRS varied by state. However, in most states this program significantly reduced UI receipt, especially for those at higher risk of exhaustion. Effects seem to vary across states depending on the intensity of services that were provided. There were no consistent impacts on employment or earnings.
Friedlander and Burtless (1995)	W2W	Varied	Welfare women	1) Tests four W2W programs operated under WIN. Most emphasized job search assistance, some also provided paid work experience where job search was unsuccessful. However, two also provided education and training. 2) All programs caused participants to find jobs more quickly. Mostly for this reason, they also caused increased earnings. These earnings effects persisted for some time, usu. a couple of years, but then decayed. Only in Baltimore (which provided some skills training) did effects persist; only in this program were effects on wages in evidence.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Friedlander and Gueron (1990)	W2W	Varied	Welfare women	1) Low cost job search programs seem to produce a lot of short term gain and are very cost-effective. 2) More intensive interventions may produce greater earnings gains that last longer, but are much more expensive. 3) Overall, if dollars are scarce, the low-cost approach seems to be better in the aggregate. 4) Widespread low-cost services coupled with careful targeting of more intensive services to long-term welfare recipients may be the optimal strategy.
Friedlander et al. (1997)	Varied voluntary programs	Varied	Mixed	1) Overall, programs have produced modest positive effects on employment and earnings for adult men and women that are roughly commensurate with the modest amounts of resources expended on them. 2) Overall, they have failed to produce positive effects for youth. 3) Considerable uncertainty remains about the kinds of training that works best and the effectiveness of training for certain demographic groups. 4) Consistently strong evidence has accumulated that govt training has been effective for adult women, and yields a positive return on investment in the short term. The average effect, while sizable, will typically not lift these families out of poverty. 5) Results for adult men are more uncertain. Many studies found weak or no effects; however, the recent JTPA experiment found effects for men about as large as for women, representing a significant break from prior studies. 6) Training programs generally seem ineffective in producing lasting earnings effects for youth. This holds for male and female youth in all program activity clusters. Despite exceptions from some studies, even Job Corps does not seem to yield consistently positive results. 7) A long-held assumption is that, the more skill building takes place, the larger and longer-lasting effects will be on earnings. However, the evidence is mixed. OJT seems to be generally effective. Results for CRT are inconsistent. JTPA found weak positive effects on earnings (but, again, not for youth). The effectiveness of CRT may depend on the relative emphasis placed on upgrading general skills as opposed to training for a specific occupation. 8) Absence of long-term follow-up is a critical problem. Some evidence suggests that earnings may persist (e.g., Couch, Zambrowski and Gordon). GAO re-analysis of the JTPA experimental data found that effects for adults continued over 5 years of follow-up, though the later-year effects were smaller than the peak effects and were not generally statistically significant. Friedlander and Burtless found that W2W effects of rapid reemployment peaked and then declined substantially by year 5, while the effect of programs that focused more on skill building tended to persist.
Gay and Borus (1980)	Mixed	Varied	Mixed	1) Estimated impacts for diverse samples are regressed on short-term outcome measures, such as whether placed and change in hours, weeks, earnings, etc. Some of the performance indicators are statistically significant. However, R-squareds are low and none is consistently related to impacts for all programs and all race/sex groups. 3) Placement at exit seems to fare the worst. Changes in wages performs a bit better.
Glazerman et al. (2002)	Varied	Varied	Mixed	Quasi-experimental methods replicate the findings from experiments only occasionally and in a way that cannot easily be predicted. However, bias is lower when the comparison group is drawn from the study itself (rather than using national data) and when it was from the same local labor market, and when pre-intervention measures of the outcomes are used. Statistical adjustments help, but propensity scores don't seem any better than simple regression adjustments.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Gottschalk (2005)	W2W	NA	Welfare recipients	Shows that working can increase welfare recipients' sense of personal efficacy.
Greenberg et al. (2003a)	Varied voluntary programs	Varied	Mixed	1) Women have by far the highest mean training effect; moreover the distribution of effects is narrowest for women. Voc CRT, OJT, and mixed CRT and workplace training are associated with large positive effects, subsidized work results in smaller but still substantial increases, while remedial training appears to be ineffective. MDTA, CETA, and JTPA all appear to be effective for women. 2) Men have a smaller mean training effect. However, effects of the training types bounce around and yield no firm conclusions. Random assignment studies produce considerably larger earnings effects than nonexperimental studies. White men appear to benefit less than nonwhites. MDTA appears to be effective, but not CETA or JTPA. 3) Youth have a negative mean effect of $92; there is a wide distribution of effects for youth, ranging from large and positive to large and negative. Although the overall effect of training for youth is close to zero, voc crt appears to have a large positive effect. Less effective for whites than nonwhites and more effective for males than females. Effects are highly sensitive to the local unemployment rate; training is less effective as unemployment increases. Effects appear to increase with program cost. 4) Effects are overall fairly small. No evidence that training is getting more effective over time (i.e., no "learning" in taking place). No evidence that more expensive programs are better. 5) Remedial training is generally not effective; voc CRT is generally always effective (though conclusions for men cannot be drawn).
Greenberg et al. (2003b)	W2W	Varied	Welfare recipients	1) A key finding is that effects of W2W programs linger for a surprisingly long time. 2) The relationship between effect and time is an inverted U. The effect peaks at about the 11th quarter and then decays, but not rapidly. Effects don't disappear entirely until about quarter 23. 3) Effects take longer to reach a peak for HCD approaches as opposed to LFA approaches. The foregone earnings for HCD means that the cumulative effects are much greater for LFA approaches. However, the human capital provided was generally quite modest, particularly when compared to what the comparison group accessed on their own.
Greenberg et al. (2004)	Varied voluntary programs	Varied	Mixed	1) A plot shows evidence that impacts persist, and, indeed, grow slightly stronger 2) Average impact was $808 in year 1, and increased on average by an additional $32 in year 2. Thereafter impacts start to decline by about $110 per year. Impacts are thus projected to reach zero about 11 years after training. 3) The decline is more rapid for men and youth; for women, there appears to be no decline.
Greenberg et al. (2005a)	Varied	Varied	Mixed	1) Nonexperimental evaluations yield similar conclusions regarding the effects of training programs for men, women, and youth. 2) Nonexperimental evaluations tend to yield somewhat smaller estimated effects than experimental evaluations, but not significantly so.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Greenberg et al. (2005b)	W2W	Varied	Welfare recipients	In a review of 27 experimental evaluations, they find that three program features are positively related to impacts: participation in job search, the use of time limits, and the use of sanctions. Evidence is mixed regarding basic ed, voc ed, and work experience, but, on balance, they appear not to add much and are certainly not worth the added cost. Participants do better in strong labor markets than weak ones. Impacts appear to start declining at about 2-3 yrs but don't decay entirely until 5-7 yrs.
Gritz and Johnson (2001)	Job Corps	Varied	Youth	Positive program impacts on earnings accrue to those who either complete a vocation or attain a GED. By contrast, those who fail to complete derive no benefit.
Grossman & Sipe (1992)	STEP	Summer employment & remediation	Disadvantaged youth	This program led to no significant long-term (approximately 4 years after enrollment) impacts on employment rates, or grades, test scores, dropout rates, college attendance, sexual behavior, teen pregnancy, or welfare receipt.
Grubb (1996)	Varied	Varied	Mixed	1) MDTA, CETA, and JTPA all show effects are most likely for females, somewhat less likely for males, and unlikely for youth. 2) W2W programs, which mostly emphasize job search assistance, also produce effects. 3) Minority Female Single Parent Demonstration, which emphasis basic skills remediation, shows no effects on earnings, except for the one (of 4) site, CET. 4) Especially disappointing results for youth programs: mixed findings re Job Corps, New Chance (which emphasized remediation) shows no effects on earnings, JOBSTART (a program like Job Corps but less intensive and without the residential component) had no effects except for the program at CET, and STEP (providing remediation and work experience for two summers for in-school youth to stem summer learning loss) also had no effect. 5) Re target populations, females benefit more than males, and adults more than youth. Unclear whether hard-to-serve benefit more or less. 6) Very quick job search services seem to provide benefits and are very cost-effective, but the effects are small; more intensive programs may also work and have larger effects, but only if designed well 7) Not a lot of evidence on whether effects last. In general, they don't seem to (W2W programs seem to have effects for a couple of years only). However, there are exceptions, such as CET and the Baltimore W2W program. 8) There is a lot of inter-site variation in effectiveness for programs of the same type. 9) Various explanations for why effects aren't greater: the interventions are modest, services may be of poor quality, there are no continuing supports, or target population is in some sense hopeless (they may have poor motivation or abilities, which explains why they didn't do well in the first place).
Hamilton and Friedlander (1989)	W2W	Work-first w/ some training	Welfare women	1) SWIM operated in San Diego under WIN from 1985 to 1987 (pre-GAIN). Consisted of a fixed sequence of activities—two weeks of job search workshop, followed by three months of unpaid work experience coupled with biweekly job club sessions, followed by education and training for those still unemployed. This sequence is in contrast to GAIN, which starts off with basic education for those determined to need it. 2) Among AFDC registrants, SWIM produced earnings gains and welfare savings that were among the highest of those found in similar evaluations elsewhere. SWIM was cost effective, but the program had little effect on net income for participants

Appendix B:
Annotations

Report	Program	Service	Target	Results
Hamilton et al. (2001)	W2W	LFA HCD	Welfare women	1) Used a 5-yr follow-up period and studied 11 W2W programs. Nearly all increased the number of quarters of work and earnings, and all decreased welfare receipt. However, combined income was largely unaffected. 2) Human capital (HCD) approaches did not produce added benefits; in fact, labor force attachment (LFA) approaches generally had larger effects on employment, earnings, and welfare receipt, moved people into jobs more quickly, and were much cheaper to operate. 3) Programs with low enforcement had the smallest effects. 4) Portland had the largest effect. It emphasized waiting for a good job, not taking just any job.
Hamilton (2002)	W2W	Varied	Welfare women	1) Most people on welfare find jobs on their own (move in and out of the labor force). 2) All the W2W programs increased earnings, but the effects taper off by 5 yrs (they persist in only 2 of the 11 programs). However, effects on decreased welfare recipiency seem to persist. 3) Employment programs (LFA) had larger effects on employment and earnings than did HCD programs; both program types were equally effective in reducing welfare. 4) HCD programs might work if participants receive a high dosage, complete the training, and receive a certificate at the end. In actuality, too few people achieved these things in these programs.
Heckman, LaLonde, Smith (1999)	Varied	Varied	Mixed	1) Collectively, U.S. experimental evaluations provide some compelling evidence that the opportunity to receive these services sometimes can improve participants' employment prospects and that the resources spent on these services can pass a standard cost-benefit test. The most consistent evidence in this regard is found for adult women. For this group, effects are modest in size (ranging in size from a few hundred dollars to a thousand dollars annually), often persist at least for several years without signs of decay, arise from a variety of treatments, and sometimes appear to be remarkably cost effective. More expensive WE and training programs result in larger absolute earnings gains. Because of substantial dropout rates among the treatment group and substitution effects among the control group, the impact of services on those who actually receive them is even larger. 2) The experimental results for youth are not encouraging. 3) Effects seem to come about because of increases in employment rather than wages. 4) Results for specific programs often vary greatly by demographic group (e.g., females vs. males, adults vs. youth). Impacts also often greatly vary across program sites in the same study. 5) Non-experimental evaluations generally reinforce the findings from experimental studies. This is so despite concern about non-experimental methods' inability to control for selectivity bias. However, the variability of estimated program impacts is substantially larger, with differences due to how the matching is done, what data sources are used, etc. And modest differences in estimated impacts can have dramatic effects on calculations on the net social benefit of government programs. 6) Much less is known about impacts for dislocated workers. However, what little is known suggests that JSA is a cost-effective service for displaced workers, as is also true for economically disadvantaged adults. Second, CT and OJT confer only modest or no additional benefit. 7) There is no evaluation method of choice—all have pros and cons. Non-experimental methods can be effective in eliminating bias from estimates if comparable people are compared, they are in the same labor market, their employment histories are taken into account, and similar data collection is undertaken for them

Appendix B:
Annotations

Report	Program	Service	Target	Results
Heckman, Hohmann, and Smith (2000)	JTPA	CRT	Mixed	1) Only looks at CRT, and notes that a large fraction of treatment group members dropped out before undertaking training—only about 49% of adult males received CRT to a high of 59% of female youth. 2) Conversely, a large fraction of control group members received training from other sources—27% of adult males did so and almost 40% of female youth did. Often, this was training paid for by other public programs, such as Pell and Perkins 3) Thus, the gap in receipt of CRT between treatment and controls was a mere 20 percentage points or so. 4) Moreover, the intensity and quality of classroom training received by the two groups seems to be quite similar (though control group members seemed to receive training over a longer period of time and didn't start as soon). 5) In the limit, if all programs are perfect substitutes, no program will be shown to be effective. Thus, the dropout and substitution that occurred could give a highly misleading impression. 6) Used nonexperimental methods to develop estimates of the effect of training, whether or not received from JTPA. These estimates are very variable, depending on exactly what assumptions are made. However, they are generally all positive and are well in excess of the experimental estimates of the effect of JTPA.
Heckman, Heinrich, Smith (2002)	JTPA	Mixed	Mixed	1) To the extent that the short-term measures don't relate well to long-term impacts, the use of performance standards could actually misdirect activity by focusing program administrators' attention on criteria that may not be related to long-run net benefits—or could indeed be perversely related to them. Cream skimming is potentially one such distortion. However, this undermines economic efficiency only to the extent that participants passed over would have better long-term gains from participation than participants actually selected. 2) They find that impacts don't differ much across client characteristics. This suggests little efficiency gain or loss from creaming. 3) There are many negative relationships between short-run performance indicators and the experimental impact estimates. Moreover, R-squareds are all very low. Things are no better when you use post-program (13-week) measures. 4) Overall, JTPA performance standards do not promote efficiency because the short-term outcomes they rely on have essentially a zero correlation with the long-term impacts of employment and earnings.
Hollister (1990)	MFSPD	Varied	Minority single mothers	1) The MFSP experiment suggests that integrating remedial education with skills training at the same site is very promising. 2) Many of those assessed as needing remediation failed only the math portion. 3) Only a third of those assessed as needing remediation actually took it within four months, and those who did take it had low completion rates.
Hotz, Imbens, Klerman (2000)	W2W	Varied	Welfare women	Some prior W2W results are misleading--it is hard to estimate the effect of various types of treatments, because each treatment often varies by site (not within site). Moreover, short-term evaluation results can be misleading. The relative ranking of programs varies over time in a way that is consistent with a prior expectations.
Jacobson & Petta (2000)	Labor Exchange	Job referrals	Unemployed	Shows substantial effects on the duration of unemployment for those who received ES job referrals.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Jacobson et al. (2002)	NA	Community college	Displaced workers	1) Impacts of community college training are about 3-4 percent of earnings, or about $20 per completed credit. 2) Impacts are about the same for men and women, though minority men benefit less. Impacts are also larger for those who are more experienced and better educated to begin with. 3) Impacts were almost entirely due to large impacts associated with the health field, in technically-oriented vocations, and in academic math and science.
Jacobson et al. (2003)	NA	Community college	Displaced workers	Older displaced workers reap about the same per-period earnings boost from community college retraining as younger displaced workers, suggesting that retraining is a reasonable choice for both groups. However, returns are substantially greater for some fields of study than others. Moreover, these results should not be used to suggest that retraining is a sound social investment, particularly for older workers.
Jacobson et al. (2005)	NA	Community college	Displaced workers	Notes that displaced workers experience substantial earnings losses. As shown in the Worker Adjustment Demonstrations, job search assistance can help with the readjustment, but evidence on the impact of retraining is weak. Community college training can help substantially with earnings replacement, at least if the course of study is carefully chosen. However, this training is costly and may not be a prudent social investment.
Johnson et al. (1985)	Labor Exchange	Job referrals	Unemployed	Receiving an ES job referral results in increased earnings and a reduced duration of unemployment for women, but no effects are found among men.
King (2004)	Varied	Varied	Mixed	Reviews the literature for disadvantaged adults and youth, dislocated workers, and welfare recipients. 1) Adult women fare better than men. Youth show little effects, but Job Corps may be an exception. 2) Among DWs, job search seems to be effective, but only for the short-term, while training seems ineffective. 3) Job search seems to work for welfare recipients in the short-term, but the effects dissipate; training may help more in the longer run.
Kogan et al. (1991)	JTPA	Varied	Mixed	1) Studied a wide variety of different training programs in 15 randomly selected SDAs. 2) Quality of classroom training appears strong along many dimensions (occupational relevance is high, use is made of practical exercises, etc.). However, some programs are exceptions. 3) OJT was of mixed quality, with 30% of the ones studied rated as poor and 45% rates as good. However, the OJT payment to employers usu. did not seem to increase the customers' access to training beyond what the customer would likely have accessed anyhow.
Krueger (2003)	Varied	Varied	Mixed	Very promising results for Job Corps. Even JTPA results for youth, taken as so discouraging, are equivocal. This is because: control group kids got very equivalent services elsewhere, the follow-up period was very short (and Job Corps shows that returns don't kick in until 2 years after random assignment, due to the steep age-earnings trajectory for youth). Furthermore, the GAO findings using longer follow-up showed consistent positive differences in favor of those in the treatment group.

Appendix B:
Annotations

Report	Program	Service	Target	Results
LaLonde (1995)	Varied voluntary programs	Varied	Mixed	1) "'The best summary of the evidence about the impact of past programs is that we got what we paid for." Public investments are exceedingly modest compared to the magnitude of the skill deficiencies that policy makers are trying to address. 2) An additional year of schooling is worth about an 8% increase in earnings. Thus, this is a reasonable expectation for something that intensive. Public E&T programs are usu. much less intensive, so smaller increases in earnings should be expected. 3) Experimental and nonexperimental results generally agree—programs have generally a small effect at best, in keeping with the low intensity of the intervention. At best, earnings are increased by about $1,000 to $2,000 per year. However , little is known about how long-lasting these effects are. 4) Gains are most consistently found for adult women. Often, there are no effects for men and youth. 5) For women, findings show that gains exist but are modest, persist for at least several years, and are sometimes achieved at very little cost. Strategies that provide only low-cost job search assistance sometimes can significantly raise adult women's post-program earnings, as in the WIN experiments and work-welfare demonstrations. Is most effective when it teaches occ skills rather than basic skills. 8) For men, JTPA raises adult men's earnings modestly if they were in OJT. 9) Less favorable results for youth, including JTPA, JOBSTART, and NSWD. 10) Less is known about the effectiveness of programs for dislocated workers. However, job search assistance appears to be cost-effective. CRT or OJT appears to add little additional benefit. As with programs for disadvantaged adults, females usu. benefit more than males. 11) Overall, training may lead to earnings gains, but they are not large enough to lift people out of poverty. This finding should not be surprising given the small investments being made (in fact, the programs would need an extraordinarily high rate of return for anything else to be the case).
Leigh (2000)	Varied	Varied	DWs	1) A large volume of evidence supports the effectiveness of job search assistance for dislocated workers. 2) Reemployment bonuses seem to work less well. Although they help reduce the duration of unemployment by a small amount (a half week), they are very costly. 3) The impact of CRT seems to vary widely, depending on the study. However, in at least some studies, wage impacts seem to persist for the longer-term. One problem is that many dislocated workers are reluctant to undertake training. 4) Wage subsidy (OJT) programs also seem to have strong payoffs, even larger than for CRT. 5) There is very little interest in self-employment.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Lerman (2000)	Varied voluntary programs	Varied	Youth	1) JTPA shows no positive effects on earnings for youth, through 5 yrs after randomization. Possible reasons for these poor results: while treatment group members were getting training, control group members were getting work experience. 2) Job Corps quasi-experimental results show that Job Corps raised the rate of high school GED, reduced arrest rates, and increased earnings. 3) JOBSTART substantially increased the receipt of education, training, and employment services, and raised the pct with a GED, but had no effect on earnings. 4) Job creation programs, such as NSWD, also don't seem to produce earnings impacts. 5) Why don't they benefit? Possibly because youth lack the foundation skills to benefit from additional training. 6) Programs that do seem to work are CET, STRIVE, and YouthBuild--linking voc and basic skills, intensive case management, ensuring youth are committed to succeeding. However, the latter two have not been subject to rigorous impact evaluations of employment outcomes.
Lillard and Tan (1992)	Private sector trng	Varied	Incumbent worker	1) Private sector training taken in the current period causes earnings to drop, unless it is company training, but increases earnings in subsequent years. 2) Company training has the largest effect on earnings, followed by training from business and vocational schools. 3) However, the effect of training decays over time, by about 1.1% for each year that elapses since the training occurred. This effect is roughly constant regardless of the source of the training. 4) Given the larger initial effect of company training, its effects persist the longest (about 13 yrs), followed by business/technical school (9 yrs), and regular school (8 yrs). 5) Effects are roughly similar on unemployment probabilities.
Mallar et al.(1982)	Job Corps	Varied	Youth	Participation in Job Corps appears to markedly increase earnings in the period just after youth leave the program, and the effects appear to persist during the four-year postprogram period for which these authors have data. Program completers appear to benefit the most, while those who drop out early benefit little or not at all.
McClanahan et al. (2004)	Summer Career Exploration Program	Work experience	Teens	The program dramatically increased access to summer work experience in private-sector jobs. However, this yielded no impacts on subsequent employment (during the next school year), future plans or intentions, college enrollment, sense of self-efficacy, or criminal activity.
Meyer (1995)	Labor Exchange	Job search & reemployment bonuses	Unemployed	1) Two main forms of UI experiments: re-employment bonuses (there are four of these) and job search programs (six of these). 2) Bonuses do speed re-employment somewhat but are costly and the effects are small. Also, they could have the unintended consequence of causing the temporarily unemployed to file for UI when they wouldn't have bothered otherwise. 3) Job search programs seem to be virtually uniformly successful in slightly reducing unemployment at very low cost, and don't seem to decrease--and may even increase--earnings in the short run, suggesting that workers are not being led to accept lower paying jobs rather than better jobs that more protracted job search would have led them to. However, effects are difficult to disentangle, because many of the experiments also imposed increased reporting requirements on participants.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Miller et al. (2003)	CET Replication	Integrated basic & occ skills instruction	Youth	1) DOL funded 12 sites to replicate the CET model in the early 1990s. Served youth 16-22. 2) Implementing CET is difficult. Only 4 sites implemented the model well. 3) In these four sites, access produced positive impacts on the probability of being employed for young women, and their earnings were higher (but not significantly so) 30 months after randomization. For young men in these sites, results were either negative or negligible. 4) In medium and low fidelity sites, impacts were either negative or negligible across a range of outcomes. 5) These results are similar to CET San Jose results, where positive impacts are driven by results for young women. 6) Gap in service receipt between treatments and controls was low on average, but this varied by site.
Nudelman (2000)	JTPA	Varied	Welfare women	1) Gains in earnings were evident. Were largest in the first year following treatment, at about $1,200 per enrollee. Were still significant in the second year, but smaller 2) Largest effects were for OJT. In contrast, CRT was not significant, possibly because CRT was generally short-term.
O'Leary (2004)	Labor Exchange	Labor exchange and others	Unemployed	1) Job search assistance programs are very cost-effective; large-scale public service employment programs are the least effective and most costly; job training programs and employment subsidies are in between. 2) Job referrals were found to be effective in increasing earnings and reducing the duration of unemployment among females in the National Evaluation of ES (Johnson 1983); there were no effects for males. The NCES Study for Dislocated Workers in PA (Katz, 1991) also finds that ES appears to be effective and that people turn to ES when they have been unsuccessful in trying other things, and that ES does help them. More recently, the Effectiveness of Referrals and Placements in WA and OR (Jacobsen and Petta) finds that referrals seem to appreciably reduce duration of joblessness. 3) Job search assistance was tested in the Charleston Claimant Placement and Work Test, which had three treatment groups: a strengthened work test (e.g., with monitoring and sanctions), a strengthened work test plus placement services, a strengthened work test plus placement services plus a three-hour job search workshop. The strengthened work test had the greatest effect and was extremely cost effective. The WA Alternative Work Search Experiment used the standard work test (3 employer contacts per week plus an eligibility review interview 3 months) as the control group. The three treatment groups were: a complete relaxation of the work test, added individualized work search requirements, and the above plus a two-day job search workshop. Results show that dropping the work search requirement worked poorly. The additional requirements had some added benefits but the benefits occurred just before the individual assistance was to be provided, not after. This suggests customers stopped claiming UI to avoid having to meet the extra requirements. The Maryland UI Work Search Experiment confirms the above. 4) Profiling can be very helpful, as demonstrated in the New Jersey UI Reemployment Experiment of Dislocated UI Claimants, and the Job Search Assistance Experiment. 5) Reemployment bonuses do reduce the duration of unemployment slightly (by about a half week) without inducing the unemployed to accept jobs that pay lower wages. However, from the standpoint of the UI system, they are generally not cost-effective (though much depends on how the bonus is structured). Moreover, there are potential entry effects and displacement effects. Narrowly targeted programs might obviate these problems and could have more favorable cost-benefit ratios.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Plimpton and Nightingale (2000)	Varied	Varied	Welfare women	1) Most welfare employment programs that offer job search assistance has positive impacts, but they are small and don't persist. 2) Work experience programs don't work unless they are associated with a training component, such as the NSWD and the Homemaker Health Aide Demonstration, in which case they have strong and persistent effects. However, these programs are very costly. 3) Basic skills training doesn't seem to provide much earnings gain. 4) Occ CRT works better. Most promising is the CET model (as in the Minority Female Single Parent Demonstration), which has strong effects that persist (with some decay) for at least 5 yrs. 5) Overall, effects erode after a couple of years for less effective programs, but persist for 5 yrs or more for more comprehensive programs. 6) There are greater effects on hours worked than on wages. 7) Even those interventions with the greatest impacts don't raise families out of poverty.
Polit et al. (1985)	Project Redirection	Comprehensive	Low-income teen parents	Redirection participants showed gains in schooling, employment, and pregnancy preventions 12 months after enrollment. By 24 months after enrollment, however, these gains had largely disappeared, as comparison group members caught up to their treatment group counterparts.
Quint et al. (1997)	New Chance	Comprehensive	Teen mothers who are dropouts on AFDC	1) Experimental group members received more services in greater quantity than control group members, but the differential was surprisingly modest. 2) With respect to outcomes, experimental and control group members both advanced in many ways, but experimental group members did not advance further in most respects. New Chance did boost participants' levels of GED receipt above the control group, but showed no impact on receipt of skills credentials, or on employment or earnings, welfare receipt or subsequent childbearing. This is so despite the fact that experimental sites were offering services that were judged to be high quality, and that participants liked the program and reported that it benefited them.
Riccio et al. (1994)	W2W	Remediation; job search assistance	Welfare women and UPs	1) In this study of GAIN, average impacts on earnings in year 3 were $636. Moreover, they were flat between year 2 and 3, suggesting some longer-term persistence. Also produced welfare savings Impacts were by far the largest in Riverside, the county that most emphasized quick job search and an "any job" mentality. 2) Alameda was the county that emphasized skills training. It had smaller effects and was also very costly to operate. 3) Despite these impacts, the majority of experimentals were not working at all in year 3 4) Concludes that basic ed may help explain why GAIN showed larger effects than other W2W programs, but this is by no means a sure-fire strategy, and may not be the most productive approach Instead, an emphasis on quick employment seems very important.
Roder & Scrivner (2005)	Bridges to Work	Transportation assistance, job search assistance	Inner-city job seekers	This program provides transportation service and job search assistance to help inner-city job seekers access jobs in the suburbs. No effects on earnings were found 18 months after randomization, although program participants were more likely to be employed in jobs that offered health benefits.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Schirm and Rodgriquez-Planas (2004)	Quantum Opportunity	Comprehensive after-school program	Entering high-school students with low grades	1) QOP did not increase the likelihood of graduating high school with a diploma or earning a GED, nor did it improve high school grades or achievement test scores or reduce the rate at which youth engage in most risky behaviors. However, it did slightly increase the rate of post-secondary attendance. 2) In general, nowhere near the intensity of services was provided that was originally envisioned.
Schochet, Burghardt, and Glazerman (2001)	Job Corps	Comprehensive	Youth	1) Job Corps substantially increased the education and training that participants received. More participated in education and training (93% vs. 72%), and average hours was increased by 998 hrs. This impact corresponds to roughly 1 additional yr of education. Job Corps increased receipt of GED and vocational certificates but led to slight reductions in high school diplomas. It had no effect on college attendance. 2) As expected, the earnings of the control group were larger in the early follow-up period. It took about two years from random assignment for the earnings of those in the program group to catch up. 3) Job Corps generated positive earnings beginning in the third year after random assignment, grew during the third year, and remained fairly constant through the end of the 4th year. 4) Earnings gains were due to a combination of greater hours of work and higher earnings per hour. Moreover, experimentals were slightly more likely to be employed in jobs with benefits. 5) Earnings gains were broadly similar across most key subgroups, including those especially at risk (very young, females with children) and those at lower risk. These gains are roughly commensurate with an additional year of schooling. 6) There were additional effects on some outcomes (e.g., arrest, public assistance), but no effect on others (drug use, health, fertility).
Schochet, McConnell, Burghardt (2003)	Job Corps	Comprehensive	Youth	1) During the survey period covered by Schochet et al. (2001), the pattern of impacts is similar using administrative data, although survey impacts are larger and are more likely to be statistically significant. 2) Based on admin data, there is no impact of Job Corps on employment and earnings in the 2.5 years after the 4-year period covered by the study above. Estimated impacts are now all near zero. 3) However, earnings impact for those ages 20-24 at intake or with a hs credential seem to have persisted.
USGAO (1996)	JTPA	Mixed	Adult men and women, and youth men and women	1) Finds no significant effect of JTPA on earnings or employment rates after 5 years for any group—adult men, adult women, male youth, and female youth. 2) By the 5th year, the treatment group for each of these four subgroups showed generally higher levels of employment and earnings than those in the control group, but none of the differences was significant. 3) JTPA showed some positive effects (on both employment and earnings) of JTPA for adult women and men at 18 and 30 months, but no significant effects for any year for male and female youth. Moreover, whether treatment group youth have higher or lower point estimates varies from year to year (no clear pattern). 4) DOL in its comments points out that: point estimates are higher for the treatment group in year 5 for all groups, point estimates are about as large as in earlier years (thus, lack of significance stems from higher standard errors), and these are per-assignee, not per-participant, effects.

Appendix B:
Annotations

Report	Program	Service	Target	Results
Westat (nd)	JTPA	Mixed	Adult women	Results for Adult Women: 1) Impacts are positive for the 7 yrs after randomization, but are only significant for the first four years. Effects do appear to decay. 2) Impacts are smaller and not significant in any year for those recommended for CRT. Impacts are larger and significant in some years for those recommended for OJT or Other Services Results for Adult Men. 1) Findings are not significant in any year and point estimates are fairly small. 2) Of the treatment types, OJT seems to have the largest impact (but it is still not significant). Other Services has large negative effects, while CRT bounces around. Results for Female Youth: 1) Impacts are positive in most years, but are never statistically significant. Nor is there a clear pattern by service activity. Results for Male Youth. 1) Impacts are usu. negative in most year, sometimes significantly so. OJT seems to fare best, but is never significant.
Zambrowski & Gordon (1993)	MFSPD	Mixed	Minority single mothers	In contrast to the other grantees operating MFSP, the one in CET continued to record earnings impacts for its participants even five years after randomization, but only for those with 12 or more years of schooling.